CHEMISTRY TO BIOLOGY AND BACK AGAIN:

SMALL MOLECULE REGULATION OF HIF TRANSCRIPTION FACTORS AND THE DEVELOPMENT OF A PLATFORM FOR THE DISCOVERY OF NEW BIOCATALYZED ORGANIC TRANSFORMATIONS

APPROVED BY SUPERVISORY COMMITTEE

John MacMillan, PhD
Uttam Tambar, PhD
Joseph Ready, PhD
Kevin Gardner, PhD

TO MY FAMILY.

CHEMISTRY TO BIOLOGY AND BACK AGAIN: SMALL MOLECULE REGULATION OF HIF TRANSCRIPTION FACTORS AND THE DEVELOPMENT OF A PLATFORM FOR THE DISCOVERY OF NEW BIOCATALYZED ORGANIC TRANSFORMATIONS

by

Jamie Lee Rogers, B.A.

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

May, 2013

REGULATION OF THE HYPOXIC RESPONSE PATHWAY VIA SMALL MOLECULE INHIBITORS: THE DEVELOPMENT OF A NEW PLATFORM FOR THE DISCOVERY OF NOVEL BIOCATALYZED ORGANIC TRANSFORMATIONS.

Publication No._____

Jamie Lee Rogers, Ph.D. (with Degree)

The University of Texas Southwestern Medical Center at Dallas, 2013

Supervising Professor: John B. MacMillan, Ph.D.

This work is divided into two parts. The first is the description of the regulation of the hypoxic response pathway via small molecule inhibitors. The hypoxia response pathway is a way in which cells sense and regulate oxygen levels in cells. Specifically, when oxygen levels in the cells are low, a family of transcription factors known as hypoxia inducible factors (HIFs) is up regulated. Importantly, the hypoxic response pathway is often mis–regulated in cancer. As a result, the regulation of this pathway offers a promising target for cancer treatment.

Previous work on HIF identified a large internal cavity, which provided an opportunity for allosteric binding and regulation of HIF. After identification of a small molecule inhibitor of

HIF through a high–throughput screening campaign, an SAR analysis was performed on the lead molecule. This led to a greater understanding of the structural requirements to strong binding with HIF. In addition, a sterol natural product was identified through the screen that also inhibits HIF. This molecule led to the search for the endogenous ligand of the HIF transcription factor and has developed a better understanding of the natural regulation of the hypoxic response pathway.

The second part of this work describes the development of a new discovery platform for the identification of new, biocatalyzed organic transformations. Biocatalyzed organic transformations have been used by organic chemists for decades as these reactions offer many benefits to the synthetic chemist. For example, reactions catalyzed by biological enzymes tend to be very stereo- and regiospecific. In addition, biocatalyzed transformation can occur on un-functionalized organic substrates. However, research into new biocatalyzed reactions has been limited due to the challenges in searching for these new reactions. We have developed a discovery platform designed to screen a vast library of bacteria for new reactivity by introducing ^{13}C labeled substrates to cultures. This platform is illustrated with the discovery of a tunable indole oxidation reaction.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

"Development of inhibitors of the PAS-B domain of the HIF-2α transcription factor." Rogers, J. L., Bayeh, L., Scheuermann, T., Longgood, J., Caldwell, C., Key, J., Naidoo, J., Melito, L, Shokri, C., Frantz, D., Bruick, R., Gardner, K., MacMillan, J., Tambar, U., *J. Med. Chem.* **2013**, *56*, 1739–1747.

"A labeled substrate approach to discovery of biocatalytic reactions: A proof of concept transformation with *N*-methylindole." Rogers, J. L., MacMillan, J. B., *J. Am. Chem. Soc.* **2012**, *134*, 12378–12381.

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

LIST OF DEFINITIONS

Abbreviation	Definition
HIF	hypoxia inducible factor
VEGF	vascular endothelial growth factor
EPO	erythropoietin
TGF-α	transforming growth factor alpha
LOXL2	lysyl oxidase homolog 2
mTOR	mammalian target of rapamycin
PI3K	phosphoinositide 3-kinase
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
PAS	per-ARNT-sim
ARNT	aryl hydrocarbon receptor nuclear translocator
HSQC	heteronuclear single quantum coherence
NMR	nuclear magnetic resonance
SAR	structure activity relationship
DMF	dimethyl formamide
THF	tetrahydrofuran
TsOH	*p*-toluenesulfonic acid

Ac	acyl
RP-HPLC	reversed phase high performance liquid chromatography
COSY	correlation spectroscopy
HMBC	heteronuclear multiple bond correlation
TOCSY	total correlated spectroscopy
IC_{50}	½ maximal inhibitory concentration
DMSO	dimethyl sulfoxide
TBS	*tert*-butyldimethylsilane
LovD	Leiden Open Variation Database
DHEA	dehydroepiandrosterone
LDA	lithium diisopropyl amide
^{i}Pr	isopropyl
Bn	benzyl
Ts	tosyl
psi	pounds per square inch
IPA	isopropyl alcohol
mCPBA	*meta*-chloroperoxybenzoic acid
NADPH	nicotinamide adenine dinucleotide phosphate
BVMO	Baeyer–Villiger monooxygenase

CPMO	cyclopentanone monooxygenase
CHMO	cyclohexanone monooxygenase
^{t}Bu	*tert*-butyl
TFA	trifluoroacetic acid
acac	acetylacetonate
BME	2-mercaptoethanol
OTf	triflate
Dppf	1,1'-Bis(diphenylphosphino)ferrocene
GC–MS	gas chromatography mass spectroscopy
NIST	National institute of Standards and Technology
DCE	dichloroethene
PANI	polyaniline
PCC	pyridinium chlorochromate
DMDO	dimethyldioxirane
RPM	revolutions per minute

CHAPTER ONE

HYPOXIA INDUCIBLE FACTOR-2 (HIF-2) AS CHEMOTHERAPY TARGET FOR MULTIPLE CANCERS

1.1 The hypoxic response pathway

The human body is a complex network of interconnected biochemical pathways that work to regulate processes and respond to changes and stresses from the surrounding environment. These pathways are necessary for the normal function of the human body and allow us to adapt to changes in our surroundings. When errors occur within the regulation of these pathways, this can often result in a disease state for the individual.

1.1.1 Importance of the hypoxic response pathway to cancer cells

One such pathway that is particularly important in mammalian cells is the hypoxic response pathway. Mammalian cells rely on oxygen as a source of energy for cellular processes. When oxygen levels in cells are low (hypoxia), the hypoxic response pathway senses these changes and provides mechanisms for the cell to adapt and respond to the changes.

This pathway becomes particularly important in many cancer cells. One of the hallmarks of cancer is the uncontrolled growth of abnormal cells.[1] As such, cancer cells are more rapidly growing and dividing than typical healthy cells. This rapid growth exceeds the tissue's ability to obtain oxygen and these cells enter into a state of hypoxia. Therefore the hypoxic response pathway becomes essential as the cells adapt their metabolism to accommodate the new cellular environment while also building new blood vessels to increase oxygen flow.

Figure 1.1 illustrates some of the important players in this complex pathway.[2] At the heart of the hypoxic response pathway is a family of transcription factors named hypoxia inducible factors or HIFs. **Figure 1.1** also illustrates that the active form of HIF is responsible for the regulation of multiple genes including VEGF, EPO, TGF–α, and LOXL2; all of which are important for the survival of cells under hypoxic conditions, and important for the survival and aggressiveness of tumors. Specifically, EPO is involved in cell survival and erythropoiesis. VEGF aids in angiogenesis, the process by which cells build new blood vessels. TGF-α is responsible for cell motility and LOXL2 is involved in invasion and metastasis.[3]

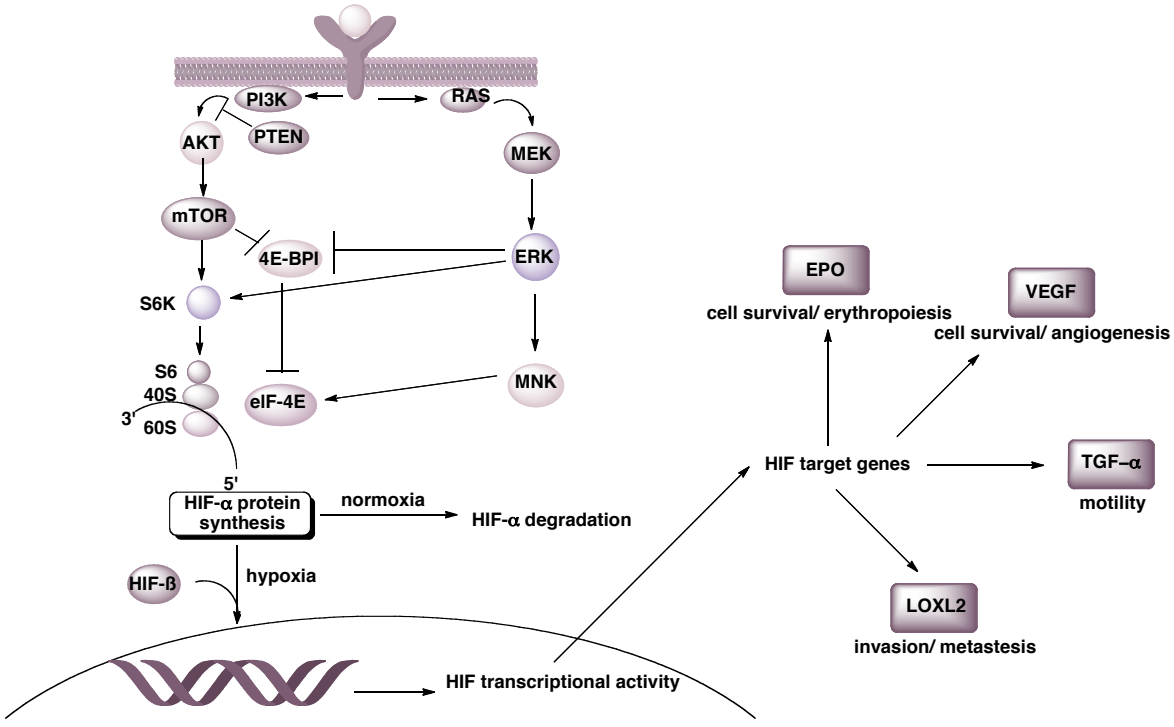


Figure 1.1 The hypoxic response pathway.

1.1.2 Current targets within the hypoxic response pathway

The hypoxic response pathway is clearly important to the survival of cancer cells, and as a result, this pathway represents an attractive target for cancer treatment. Currently, research efforts are focused on certain elements of this pathway. For example, mTOR, PI3K, and VEGF are three of the most popular targets for cancer therapeutics. In fact, 60 mTOR inhibitors are currently in preclinical or clinical trials, 90 PI3K targeting drugs are being developed and 70 VEGF inhibitors are either in clinical development or have been launched while an additional 70 are in preclinical studies.[4]

While these targets are clearly relevant to cancer therapy, there are complications to the development of chemotherapies focused on them. As **Figure 1.1** illustrates, both mTOR and PI3K are kinases that play an important role in the regulation of multiple other kinases and proteins. Targeting these kinases can have detrimental effects on healthy cells as well. mTOR for example is important to the cell cycle progression. It regulates essential signal transduction pathways, and likely plays a role in the transcription of DNA to RNA.[5] Damage to healthy cells by mTOR inhibition can be reversible, but this requires intermittent administration of the cancer drugs to allow the patient to recover between treatments. Ultimately, this decreases the efficacy of mTOR inhibitors for the patient.

VEGF is key to angiogenesis, and therefore it was thought that by blocking VEGF, the blood supply to cancer cells would be cut. This approach was thought to be a way of starving cancer cells. However, it has become clear that the mechanism of action for VEGF inhibitors is more complicated. It is more likely that multiple mechanisms are leading to the efficacy of these drugs.[6] As with other cancer treatments, targeting VEGF is not effective for all patients.

While mTOR, PI3K, and VEGF all remain valuable targets for cancer treatment; there is still a need for new drug targets. The HIF transcription factors represent a novel target for cancer chemotherapeutics that has largely been under–explored. HIF plays an important role in several processes that are key to the survival and proliferation of cancer cells. HIF regulation as a cancer therapeutic target provides a nice balance between the broadly acting mTOR and PI3K inhibitors and the more specific VEGF inhibitors. This review will focus on the progress that has been made in establishing HIF as a viable target in cancer treatments.

1.2 Indirect targeting of HIF

Currently many natural products and synthetic small molecules have been shown to indirectly inhibit the HIF transcription factors. This is often accomplished through the inhibition of HIF protein synthesis or enhanced degradation of the HIF protein.[7] For example, a semi–synthetic natural product derivative GL331 (**Figure 1.2**), was a clinical candidate that showed an inhibitory effect on the growth of several cancer cell lines including gastric, cervical and lung cancers. GL331 was shown to inhibit vascular growth within cancer cells. GL331 works through the down–regulation of HIF–1α expression.[8]

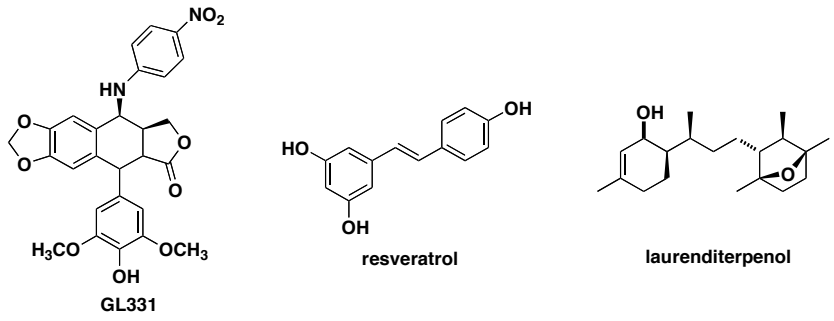


Figure 1.2 Indirect inhibitors of HIF

In another example, resveratrol, a secondary metabolite produced by a large variety of plants under environmental stress, was shown to inhibit HIF–1α protein expression through multiple mechanisms. Resveratrol blocks HIF–1α protein synthesis and also promotes the protein degradation through the proteosome.[9]

Finally, laurenditerpenol is a marine–derived secondary metabolite that inhibits HIF–1 activation under hypoxic conditions. Laurenditerpenol acts by blocking the protein accumulation that typically occurs under hypoxic conditions. In this case, mitochondrial oxygen consumption is blocked and this keeps cellular oxygen concentrations at a level that promotes HIF-1α degradation.[7,10] In each of the examples presented here, HIF activation is targeted indirectly, either by blocking initial HIF transcription or through increasing HIF degradation. There are currently no known direct inhibitors of HIF.

1.3 HIF structure

1.3.1 Dimeric structure of HIF

The hypoxic response pathway is directly mediated by the hypoxia–inducible factor or HIF. HIF is a heterodimeric transcription factor that is composed of an α and a ß subunit (**Figure 1.3**). Both of these subunits are members of the basic helix–loop–helix (bHLH)– Per-ARNT-Sim (PAS) family of transcription factors. The α subunit of HIF is directly regulated by oxygen while the ß subunit is constitutively expressed and independent of oxygen concentrations. Further, there are three isoforms of the alpha subunit: HIF-1α, HIF-2α, and HIF-3α. HIF-1 and HIF-2 are structurally similar and also share many of the same functions and regulations. However, they are not redundant and may have unique functions during development.[3] For example, HIF–1α has been shown to induce the glycolytic pathway while

HIF–2α regulates genes that are involved in tumor growth and cell cycle progression.[11,12] Both HIF–1 and HIF–2 are currently being explored as potential targets for cancer. Less is understood about HIF-3 though it likely helps to regulate the other two alpha subunits.[13] The ß subunit of HIF is also referred to as the aryl hydrocarbon receptor nuclear translocator, or ARNT. Both subunits (α and ARNT) contain two PAS domains, PAS–A and PAS–B.

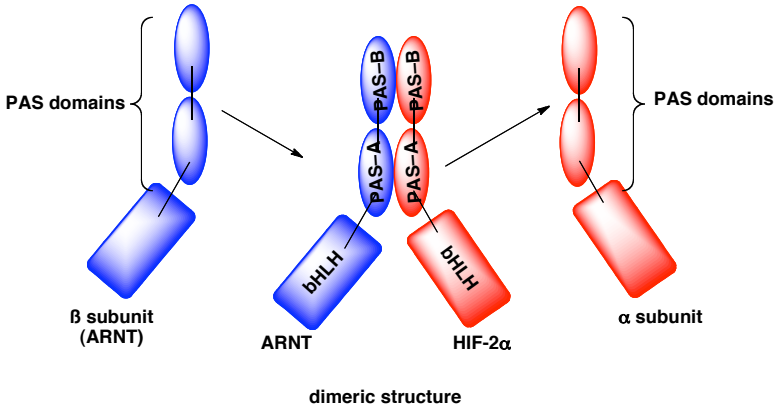


Figure 1.3 HIF–2 dimeric structure

1.3.2. Regulation of HIF dimerization

As previously mentioned, HIF is regulated by oxygen concentrations within the cell (pO_2). Under levels of high oxygen, the alpha subunit is hydroxylated at specific proline residues by a proline hydroxylase, and this labels it for ubiquitin degradation.[14] This process keeps the concentration of HIF-α in the cells relatively low. However, when oxygen levels in the cell are low, the alpha subunit is not hydroxylated and therefore it is not degraded. As the cellular concentration of the alpha subunit increases, it is transferred to the nucleus where it finds its binding partner, ARNT (**Figure 1.4**). It is the dimeric form of HIF that is then responsible for the regulation of a series of downstream targets that are essential to the survival of cells in a

hypoxic environment. The dimerization of HIF could therefore be a key to direct regulation of this transcription factor.

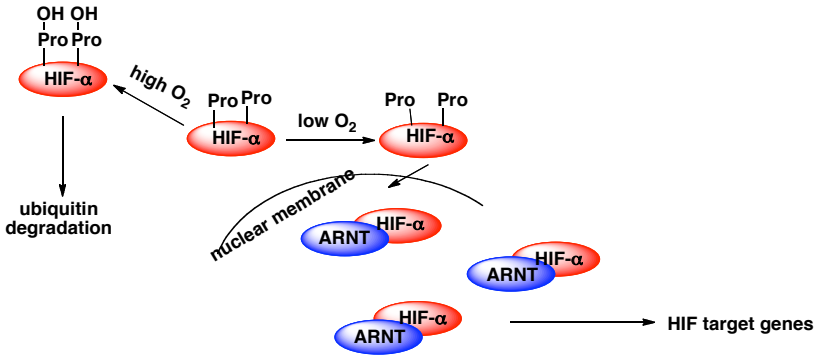


Figure1.4 Normoxia and hypoxia regulation of HIF

1.4 Factors effecting dimerization of HIF protein

1.4.1 Importance of the PAS–B domain in forming a robust dimer

To better understand the molecular requirements for this dimerization, the structure of HIF has been closely studied. The architecture of the HIF subunits is composed of two PAS domains, PAS-A and PAS-B. PAS domains are common components of signal transduction proteins[15], and are often involved in protein–protein interactions. In fact, PAS domains have been shown to promote heterodimerization of bHLH transcription factors with bHLH partner factors, necessary for DNA binding activity.[16] In 1996, the importance of the PAS domains in heterodimerization was confirmed. Immunoprecipitation experiments showed that HIF-α and ARNT exist as a heterodimer in hypoxic cells.[17] Also, using truncated versions of the HIF subunits and through coimmunoprecipition, it was shown that the PAS domains were essential to form the heterodimer that is robust enough to bind to DNA.[17,18] The bHLH domains of HIF-α

and ARNT were also able to dimerize without the PAS domains, but this interaction was weaker. To further characterize the location of the dimerization interactions, protein labeling and NMR experiments were utilized. An unlabeled ARNT PAS–B subunit was titrated into a ^{15}N–labeled HIF-2α PAS–B domain. Using $^{15}N/^{1}H$ HSQC NMR spectroscopy, chemical shift changes were observed in the spectra, which corresponded to binding interactions between the two subunits.[15] As a control experiment, the labeled HIF–2α PAS–B domain was also titrated with a PAS domain from a PAS kinase, which is not involved in hypoxia response. In this experiment, the signals from the HIF-2α PAS–B domain were not affected. The binding interactions observed in this experiment were at the solvent exposed face of the PAS domain.[15] Further, using a luciferase reporter, it was shown that interruption of the HIF dimerization also decreased the HRE–driven response.[15]

1.4.2 Location of key residues at dimer interface

The importance of this interface between the two PAS–B domains was further explored by creating point mutations of three residues on the PAS–B domain of the alpha subunit. The mutations at residues Q322E, M338E and Y342T were located at the solvent exposed face of the PAS–B domain. NMR characterization confirmed that the mutant protein retained its tertiary protein structure. However, upon titration with ARNT, only minimal chemical shift changes were reported. This suggested that the binding interaction between the two subunits had been significantly decreased.[15]

1.4.3 Internal binding cavity within PAS–B domain

Importantly, PAS domains are often regulated by ligands or cofactors through allosteric binding within hydrophobic cores.[19] As such, it was thought that the PAS domains of the HIF

transcription factor may also be allosterically regulated. Although no endogenous ligand was known, a crystal structure of the PAS–B domain of HIF–2α revealed a large (290 Å) water filled cavity (**Figure 1.5**).[20] A water–filled cavity of this size is rare and this suggests that an unknown cofactor or ligand is associated with this protein.

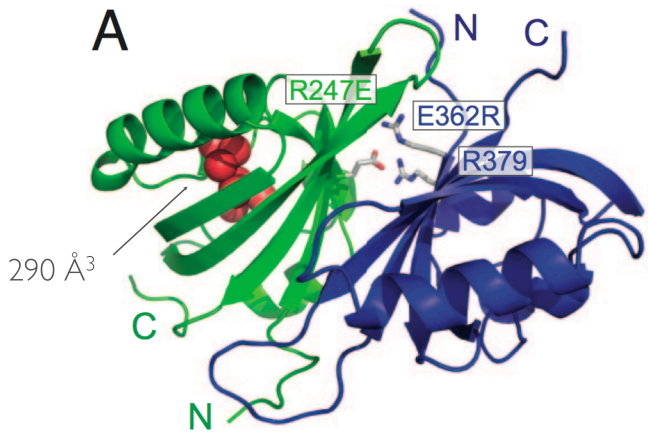


Figure 1.5 Crystal structure of HIF–2α PAS–B domain (green) and ARNT PAS–B domain. Internally bound water molecules are shown in red.[20]

1.4.4 Development of an NMR screen to identify artificial ligands.

While the identity of a natural ligand is still unknown, a lot of work has been done to identify artificial ligands for this PAS–B domain. Using the same ^{15}N labeled protein and the $^{15}N/^{1}H$ HSQC NMR spectroscopy, a series of ligands was identified that had low micromolar affinity.[21] In this screen, 750 compounds were selected that had structural features similar to many protein binding compounds such as a small molecular weight (<200 Da). Each compound was titrated with the protein and chemical shift changes evaluated. From this screen, nine compounds were identified that had equilibrium dissociation constants that were less than 100

μM. Interestingly, all of the compounds chosen showed similar interactions with the protein. This was observed by similar chemical shifts in the NMR spectra.

1.5 Development of an AlphaScreen assay for high-throughput screening of HIF modulators

A crystal structure confirmed the presence of a large binding cavity in the PAS–B domain of the HIF-2α transcription factor. Through NMR screening, it was possible to confirm that small molecules bound within this cavity with low binding affinity and were able to modulate the dimerization of HIF.[22] In order to screen a larger field of molecules, a high–throughput assay was developed using Perkin Elmer's AlphaScreen method. This assay is a homogenous bead–based luminescence proximity approach. As illustrated in **Figure 1.6**, a donor bead is attached to the HIF–2α PAS–B domain. The donor bead releases a singlet oxygen species when exposed to 680 nm light. This is received by the acceptor bead which is attached to the PAS–B domain of the ARNT subunit. The acceptor bead then releases a luminescent signal, which is read by a luminometer. The singlet oxygen has a short lifespan and can travel only up to 200 nm. If the two protein subunits are separated by more than 200 nm, no luminescent signal is released. This provides a quantitative measure of interaction between the two PAS–B domains.

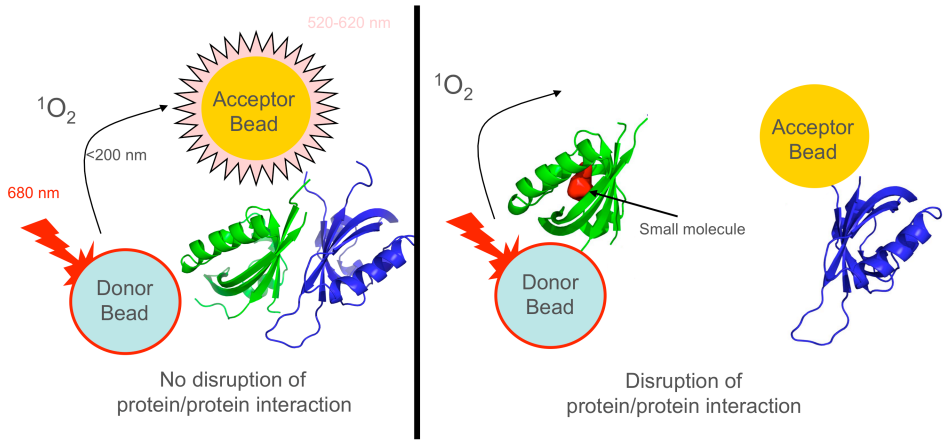


Figure 1.6 AlphaScreen for the identification of disruptors of HIF heterodimerization.

A control was also developed for this assay to help minimize the possibility of false positives. A mutation within the cavity of the PAS domain (S304M) greatly reduced binding affinity. Molecules that inhibited heterodimerization with the mutant protein were either directly interfering with the assay or inhibiting heterodimerization through some other mechanism.

With a high–throughput–screening approach in hand, >200,000 compounds from an in–house library were screened by the UT Southwestern Medical Center High Throughput Screening Facility. Of these compounds, 20 candidates warranted further analysis. The most promising compounds were similar in structure to the compounds identified in the initial NMR screen and also bound within the same cavity of the PAS domain as confirmed in NMR analysis.

In additional screening, these compounds were shown to have isothermal titration calorimetry measurments giving K_d values of 80 nM. Finally, in cell culture, it was shown that these compounds were able to interfere with transcription of downstream HIF targets with

submicromolar potency. Unfortunately, these compounds showed poor pharmacokinetic properties as shown through a decreased inhibitory effect over 24 hours.

These results were promising as they showed that small molecule ligands could be used to modulate the activity of HIF transcription factors through allosteric modulation. However, the poor pharmacokinetic properties shown by these compounds were a limiting factor. As a result, a structure activity relationship (SAR) campaign was started to identify more potent ligands with improved PK properties.

1.6 Conclusion

As one of the biggest threats to human health, cancer is a focus of many research efforts throughout the world. Chemotherapy is one of the leading forms of treatments for this class of diseases, but current chemotherapeutics are still falling short as treatments as cancer is still one of the leading causes of death worldwide. Currently, popular biological targets for cancer treatments are involved in the hypoxic response pathway. Three of the most popular targets are mTOR, PI3K, and VEGF; with nearly 300 total drug candidates in preclinical, clinical, and marketed drugs just covering these three targets. However, there are limitations to these targets as well.

HIF represents a new target in cancer therapeutics. HIF regulates over 100 genes, many of which are important to the survival and proliferation of cancer cells. In fact, up–regulation of HIF is often associated with increased tumor aggressiveness. Studies have shown that HIF dimerization is essential to its activity, and that by interrupting the dimerization, the downfield targets are impacted. Further, the discovery of an internal binding cavity within the PAS–B domain of the HIF-α subunit offers a unique target for small molecule manipulation. Studies

have shown that small molecules that bind within this internal cavity can interrupt the HIF dimerization and importantly, can turn off the downfield targets.

Our efforts in this field have focused on the development of a small molecule inhibitor of HIF that can meet the binding affinity requirements to effectively interrupt HIF dimerization. This small molecule ligand should also have improved pharmacokinetic properties.

1.7 References

1. Hanahan, D., Weinberg, R. A., *Cell*, **2011**, *144*, 646–674. Hallmarks of cancer: the next generation.

2. Semenza, G. L., *Nature Rev.* **2003**, *3*, 721–732. Targeting HIF–1 for cancer therapy.

3. Lofstedt, T., Fredlund, E., Holmquist–Mengelbier, L., Pietras, A., Ovenberger, M., Poellinger, L., Pahlman, S., *Cell Cycle*, **2007**, *8*, 919–926. Hypoxia Inducible Factor–2a in Cancer.

4. Booth, B., Cancer Drug Targets: The March of the Lemmings Retrieved: Feb 8, 2013. June 7, 2012. http://www.Forbes.com/sites/brucebooth/2012/06/07/cancer–drug–targets–the–march–of–the–lemmings/2/

5. Faivre, S., Kroemer, G., Raymond, E., *Nature Rev.* **2006**, *5*, 671–688. Current development of mTOR inhibitors as anticancer agents.

6. Ellis, L. M., Hicklin, D. J., *Nature Reviews Cancer*, **2008**, *8*, 579–591. VEGF–targeted therapy: mechanisms of anti–tumour activity

7. Nagle, D. G., Zhou, Y–D. *Curr. Drug Targets.* **2006**, *7*, 355–369. Natural product–based inhibitors of hypoxia–inducible factor–1 (HIF–1).

8. Chang, H., Shyu, K–G., Lee, C–C., Tsai, S–C., Wang, B–W., Lee, Y. H., Lin, S., *Biochem Biophys. Res. Commun.* **2003**, *302*, 95–100. GL331 inhibits HIF–1a expression in a lung cancer model.

9. Cao, Z., Fang, J., Xia, C., Shi, X., Jiang, B. H., *Clin. Cancer Res.* **2004**, *10*, 5253–5263. *trans*–3,4,5'-Trihydroxystibene inhibits hypoxia–induction factor 1a and vascular endothelial growth factor expression in human ovarian cancer cells.

10. Mohammed, K. A., Hossain, C. F., Zhang, L., Bruick, R. K., Zhou, Y. D., Nagle, D. G., *J. Nat. Prod.* **2004**, *67*, 2002–2007. Laurenditerpenol, a new diterpene from the tropical

marine alga *Laurencia intricate* that potently inhibits HIF–1– mediated hypoxic signal in breast tumor cells.

11. Hu, C. J., Wang, L. Y., Chodosh, L. A., Keith, B., Simon, M. C., *Mol. Cell Biol*. **2003**, 9361–9374. Differential roles of hypoxia–inducible factor 1 alpha (HIF–1alpha) and HIF–2alpha in hypoxic gene regulation.

12. Wang, V., Davis, D. A., Haque, M., Huang, L. E., Yarchoan, R., *Cancer Res.* **2005**. 3299–3306. Differential gene up–regulation by hypoxia–inducible factor–1alpha and hypoxia–inducible factor–2alpha in HEK293T cells.

13. Bridges, K. S., Sharp, T. V., *Future Oncol.* **2012**, *8*, 491–493. Regulators of the hypoxic response: a growing family.

14. Loboda, A., Jozkowicz, A., Dulak, J., *Vascular Pharmacology*, **2012**, *56*, 245-251. HIF–1 versus HIF–2 – Is one more important than the other?

15. Erbel, P. J. A., Card, P. B., Karakuzu, O., Bruick, R. K., Gardner, K. H., *PNAS*, **2003**, *100*, 15504–15509. Structural basis for PAS domain heterodimerization in the basic helix–loop–helix-PAS transcription factor hypoxia–inducible factor

16. Pongrantz, I., Antonsson, C., Whitelaw, M. L., Poellinger, L. *Mol Cell Biol.* **1998**, *18*, 4079–4088.

17. Jiang, B–H, Rue, E., Wang, G. L., Roe, R., Semenza, G. L., *J. Biol. Chem*. **1996**, *271*, 17771–17778. Dimerization, DNA binding, and transactivation properties of hypoxia–inducible factor 1.

18. Hogenesch, J. B., Chan, W. K., Jackiw, V. H., Brown, R. C., Gu, Y–Z., Pray–Grant, M., Perdew, G. H., Bradfield, C. A., *J. Biol. Chem*. **1997**, *272*, 8581–8593. Dimerization, DNA binding and transactivation properties of hypoxia inducible factor–1.

19. Taylor, B. L., Zhulin, I. B., *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 479–506. PAS domains: internal sensors of oxygen, redox potential, and light.

20. Scheuermann, T. H., Tomchick, D. R., Machius, M., Guo, Y., Bruick, R. K., Gardner, K. H., *PNAS*, **2009**, *106*, 450–455. Artificial ligand binding within the HIF2a PAS–B domain of the HIF2 transcription factor.

21. Amezcua, C. A., Harper, S. M., Rutter, J., Gardner, K. H. *Structure,* **2002,** *10*, 1349–1361. Structure and interactions of PAS kinase N–terminal PAS domain: model for intramolecular kinase regulation.

22. Key, J., Scheuermann, T. H., Anderson, P. C., Daggett, V., Gardner, K. H., *J. Am. Chem. Soc.* **2009**, *131*, 17647–17654. Principles of ligand binding within a completely buried cavity in HIF2a PAS–B.

CHAPTER TWO

SMALL MOLECULE INHIBITORS OF HIF–2: A STRUCTURE AND ACTIVITY RELATIONSHIP ANALYSIS

2.1 Identification of a small molecule inhibitor of HIF-2 through high- throughput screening

HIFs play an important role in regulating the expression of over 100 genes essential to the cellular response to hypoxia. They also play an important role in promoting tumor aggressiveness in multiple cancer types.[1] As such, HIF has been proposed as a therapeutic target for cancer treatment. While these transcription factors have been indirectly targeted in multiple ways,[2-5] we sought to develop a small molecule capable of directly modulating HIF.

Previous efforts by the Bruick and Gardner labs showed that disruption of the HIF heterodimer through allosteric binding of a small molecule was a feasible way to influence HIF activity. Through an in–house high–throughput–screening campaign using the AlphaScreen assay, approximately 20 candidate molecules were identified that interrupted HIF dimerization. These molecules were further validated in NMR studies that showed weak binding to the internal cavity of the HIF–2α PAS–B domain as described previously. The goal of this project was to improve upon the potency and activity of the ligand as well as its pharmacokinetic properties. Starting with compound **2-1** from the initial screen, a systematic series of changes were made to the molecule to evaluate the structure and activity relationship. Specifically, **2-1** was broken down into three regions: the A–ring, the B–ring, and the linker connecting these two pieces (**Figure 2.1**). By evaluating each region of the molecule, it was possible to determine the

structural features necessary for potent activity. In order to show the full SAR analysis performed, molecules and corresponding biological data from compounds synthesized by collaborators will be included in this report. These molecules will be clearly denoted and information about their synthesis is available.[6]

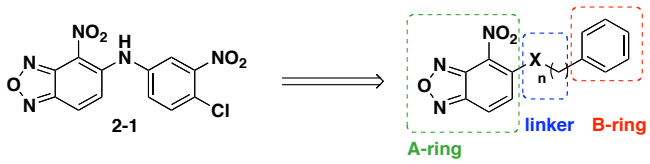


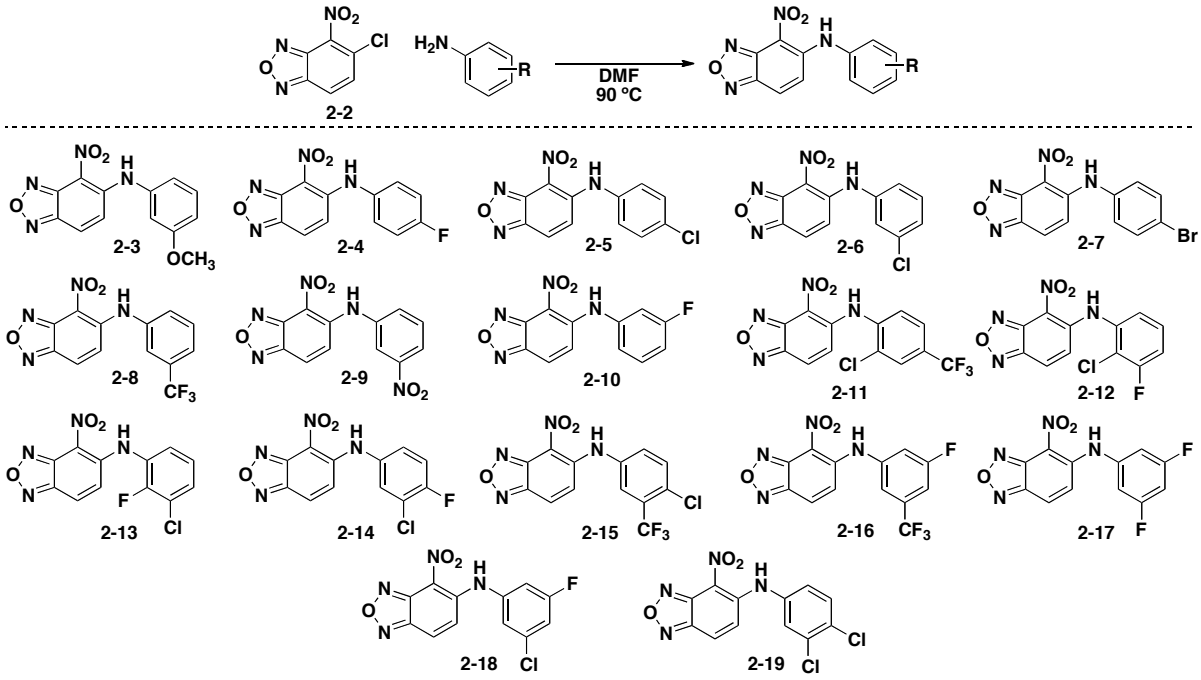
Figure 2.1 Development of SAR strategy

2.2 SAR analysis: B ring.

2.2.1 Synthesis of B–ring derivatives

Initial efforts towards evaluation of the parent compound focused on alterations to the B–ring of the molecule. Compound **2-1** is composed of two primary structural units, a benzoxadiazole ring and a substituted aniline ring, which are connected through the aniline nitrogen. To evaluate the B ring of this molecule a series of commercially available aniline rings were coupled to 4-chloro benzoxadiazole in DMF at 100 ºC (**Scheme 2.1**).

Various substitution patterns were evaluated around the aniline ring ranging from electron donating methoxy groups to an electron withdrawing nitro functionality. Substitutuents were evaluated in the ortho, para and meta positions of the B–ring. Each of these compounds was then evaluated via the AlphaScreen including the mutated protein S304M, described in the previous chapter.



Scheme 2.1 Synthesis of B-ring derivatives

2.2.2 Activity of B–ring derivatives

Clear patterns were observed for activity of the molecule with B ring alterations. The biological data for these compounds is listed in **Table 2.1**. When the biological data in **Table 2.1** is listed as ND, this signifies that the compound interrupted heterodimerization with the wild type protein, but also inhibited the S304 mutant, which suggests the molecule is interfering in some other manner (described in detail in chapter one).

Table 2.1: Biological data for compounds **2-3** – **2-19**.

Compound	AlphaScreen data (μM)	K_D (μM) ITC data	Compound	AlphaScreen data (μM)	K_D (μM) ITC data
2-3	**NA**	**–**	**2-12**	**ND**	**–**
2-4	**NA**	**–**	**2-13**	**0.76**	**1.1**
2-5	**NA**	**–**	**2-14**	**0.17**	**0.4**
2-6	**0.18**	**0.16**	**2-15**	**0.09**	**0.17**
2-7	**NA**	**–**	**2-16**	**0.43**	**0.37**
2-8	**0.46**	**0.64**	**2-17**	**2**	**0.73**
2-9	**2.8**	**3**	**2-18**	**0.12**	**0.17**
2-10	**2.1**	**2.2**	**2-19**	**0.1**	**0.09**
2-11	**NA**	**–**			

One trend observed for this series was that when the molecule was di–substituted versus mono–substituted the activity was generally higher. However, compounds **2-6** and **2-8** showed moderate activity although they were mono–substituted. The common factor for these two compounds is that they're substituted at the meta position. Meta substitution appeared to be a necessary attribute for biological activity as the di–substituted series also saw better activity when at least one substituent was located in the meta position, as observed in compounds **2-15**, **2-18**, and **2-19**. The strongest inhibitors as observed in the AlphaScreen were then evaluated by ITC. There was a strong correlation between AlphaScreen results and ITC data. From this initial analysis, compound **2-18** showed strong inhibition and strong binding. The substitution pattern for this molecule, 3–chloro–5–fluoro was carried forward into many of the derivatives in this study.

2.3 SAR analysis: new linker

2.3.1 Linker length

To evaluate the importance of the amine bridge for activity, a series of derivatives were synthesized that looked at new chain lengths and also new heteroatom substitution for this linker. The first of these looked at a series of benzylamine derivatives (*synthesized by Leila Bayeh, UTSW*), which extended the chain length from one atom (the amine) to two atoms (carbon and nitrogen) (**Figure 2.2**). These molecules were synthesized following the same protocol as the aniline derivatives.

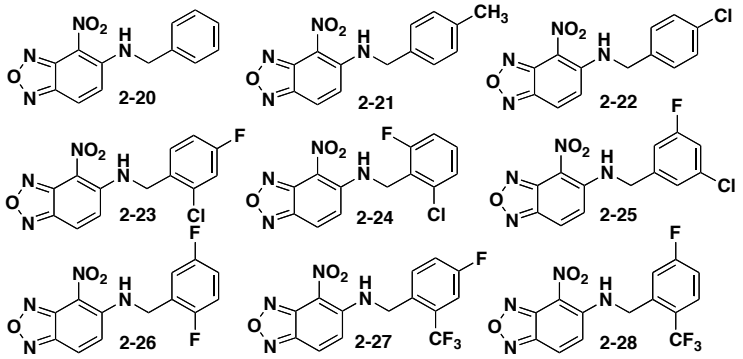
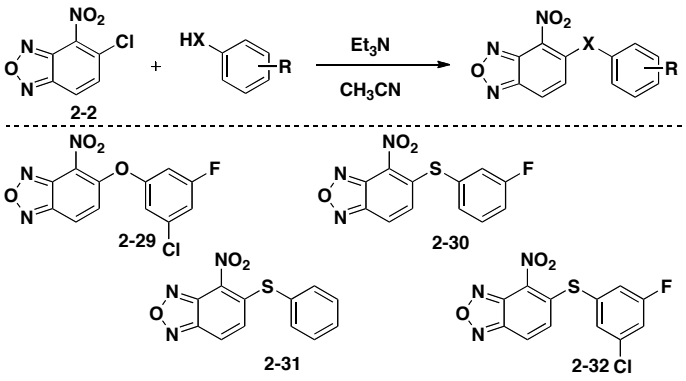


Figure 2.2 Benzylamine analogs

2.3.2 Synthesis of non–amine derivatives

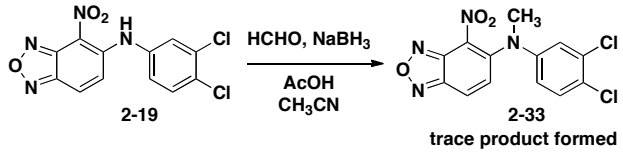
We continued to evaluate the importance of the amine linker through a series of analogs in which the secondary amine was replaced with a different heteroatom and also substituted to a tertiary amine. The first derivative replaced the amine with oxygen, **2-29** (*synthesized by Leila Bayeh, UTSW*). We also looked at a few thioether derivatives, **2-30** – **2-32** by coupling thiols with the benzoxadiazole fragment. In these coupling reactions base was required because of the lowered nucleophilicity of the thiol group (**Scheme 2.2**).



Scheme 2.2 Ether and thioether analogs

2.3.3 Synthesis of N-methyl amine derivatives

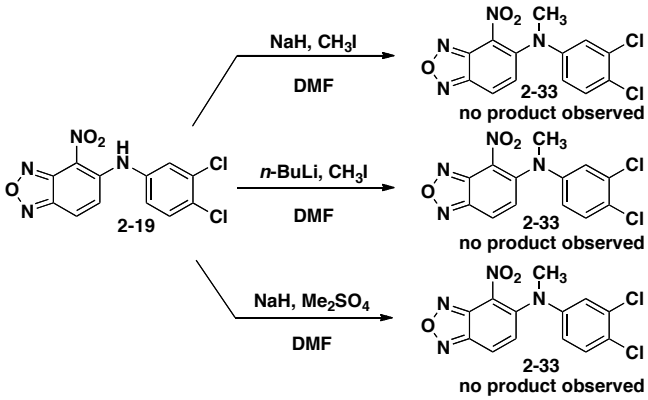
We also wanted to see what happened when the amine was substituted so we looked at the N–methylated derivative. Synthesis of this molecule was not as straight forward, however. Initially, a direct methylation reaction was attempted on the dichloro derivative (**2-19**). The first approach was using formic acid and sodium borohydride.[7] In this reaction, only trace amounts of the desired product were observed via LCMS (**Scheme 2.3**)



Scheme 2.3 Attempt at direct methylation of amine linker

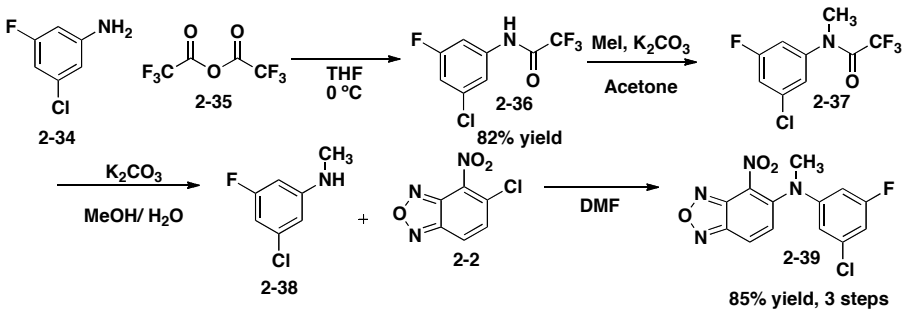
We next tried nucleophilic substitution by reacting the starting material with sodium hydride and then methyl iodide. In this reaction, no product was observed. Using a stronger

base (*n*-butyl lithium) was also unsuccessful. Finally, a stronger electrophile in Me_2SO_4 was also unsuccessful for this reaction and in each case none of the desired product was observed (**Scheme 2.4**)



Scheme 2.4 Additional efforts at direct methylation of amine linker

To obtain the N-methylated product a less direct route was used in which the un–coupled aniline was reacted with trifluoro acetic anhydride to give the N-acylated product.[8] With the amide hydrogen now significantly more acidic a weak base and methyl iodide were used to perform the nucleophilic substitution reaction.[9] Hydrolysis in the presence of potassium carbonate followed by the coupling reaction gave the desired product in good overall yield (**Scheme 2.5**).



Scheme 2.5 Successful approach to methylation of amine linker

2.3.4 Activity of linker derivatives

Each of these linker derivatives was then evaluated in the AlphaScreen. We saw diminished activity for most of the benzylamines (**2-20** – **2-28**) relative to the corresponding amines. However, two of the compounds in this series did have appreciable activity (**2-23** and **2-27**). These molecules do not have any meta substituents, a requirement for activity in the amine series. This suggests a different orientation of the B–ring with the extended linker length. Unfortunately, each of the ether and thioether derivatives interacted with the S304M mutant in the AlphaScreen, and therefore no IC_{50} was determined (**Table 2.2**). Both the ether and thioether derivatives lack the hydrogen bond potential that the lead compound has and therefore could result in decreased activity. This is supported by the methylated derivative, **2-39**, which also has no hydrogen bond donor and has no binding affinity as shown by ITC data.

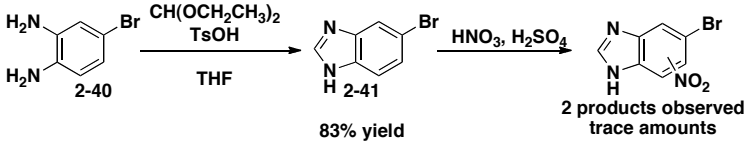
Table 2.2: Biological data for compounds **2-20** – **2-39**.

Compound	AlphaScreen data (μM)	K_D (μM) ITC data	Compound	AlphaScreen data (μM)	K_D (μM) ITC data
2-20	**>5**	**–**	**2-27**	**0.5**	0.54
2-21	**NA**	**–**	**2-28**	**NA**	**–**
2-22	**ND**	**–**	**2-29**	**ND**	**–**
2-23	**0.33**	**0.5**	**2-30**	**ND**	**–**
2-24	**NA**	**–**	**2-31**	**ND**	**–**
2-25	**2**	**2**	**2-32**	**ND**	**–**
2-26	**>10**	**–**	**2-39**	**ND**	**>10**

2.4 SAR analysis- heteroaromatic ring

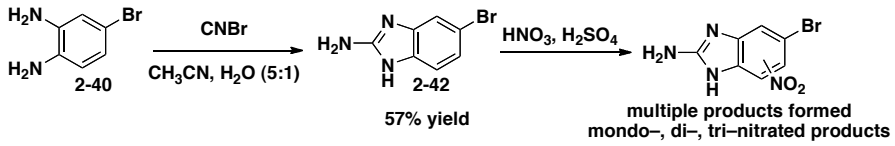
2.4.1 Synthesis of heteroaromatic ring structures

To explore the structure–activity relationship of the A-ring, a series of heteroaromatic derivatives were synthesized. Initially, the goal was to perform a direct nitration reaction on a variety of commercially available and synthetically prepared precursors (**2-41** – **2-43**). Starting from commercially available 4–bromo–1,2-diamino benzene, the 4–bromobenzimidazole was synthesized via reaction with tosic anhydride.[10] This was followed by nitration using a 1:1 ratio of nitric acid and sulfuric acid.[11] Unfortunately, two products were isolated from this reaction in only trace amounts and it was not possible to identify the position of the nitration (**Scheme 2.6**)



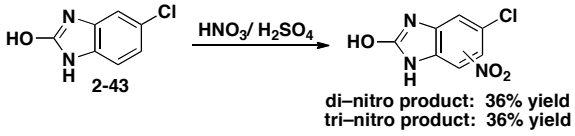
Scheme 2.6 Efforts at nitration of benzoimidazole derivative

Similary, a benzimidazole amine (**2-42**) was generated from reaction of **2-40** with cyanogen bromide[12] and again this was followed by nitration using nitric acid and sulfuric acid. In this reaction, multiple products formed and were identified as mono–, di– and tri–nitrated products (**Scheme 2.7**).



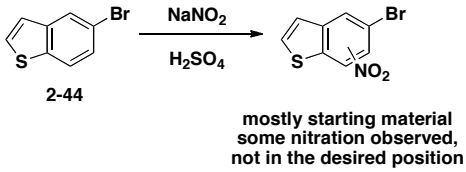
Scheme 2.7 Efforts at nitration of 2-amino benzoimidazole derivative

In another attempt at nitration, commercially available 2–hydroxy benzimidazole was nitrated under the same conditions as above and although this reaction was cleaner than the previous reactions, only a di–nitrated and tri–nitrated product were recovered (**Scheme 2.8**).



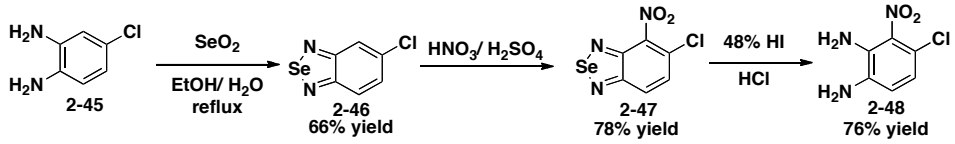
Scheme 2.8 Efforts at nitration of 2-hydroxy benzoimidazole derivative

Milder conditions were attempted for the nitration of benzothiophene. This reaction also proved difficult as primarily starting material was recovered with a small amount of nitration. However, it was not possible to determine the location of the nitration on this substrate (**Scheme 2.9**)



Scheme 2.9 Mild nitration of benzothiophene

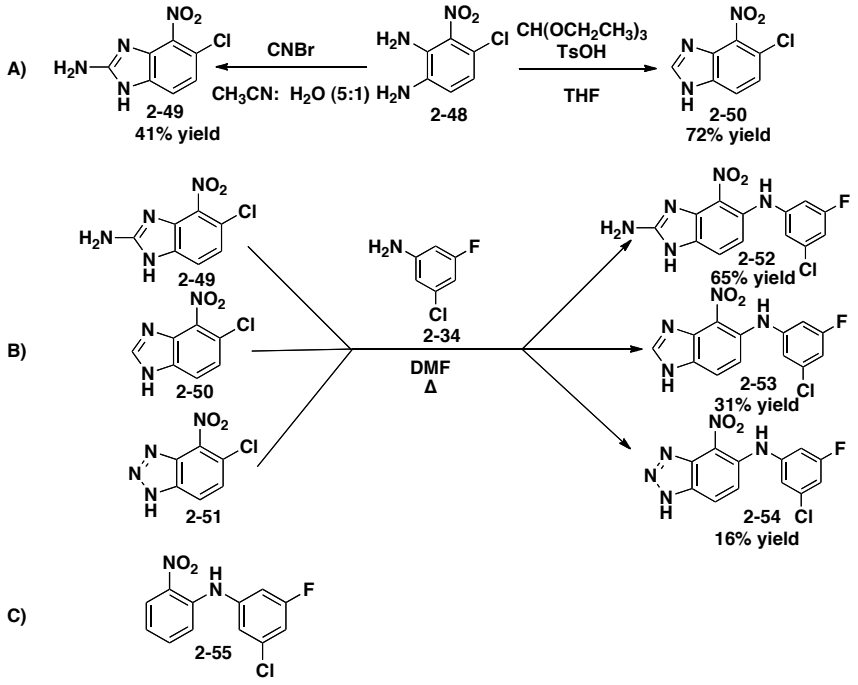
All attempts at nitration on the series of benzimidazole derivatives were unsuccessful. Separation and characterization of the nitrated products proved too difficult and an alternative approach was necessary. As an alternative approach, molecules in this series were synthesized starting from 4-chloro-1,2-diamino benzene. In order to achieve a selective and controlled nitration reaction, the substrate was first converted into the selenodiazole ring (**2-47**). This was then nitrated using standard conditions of nitric acid in sulfuric acid. 58% HI was used to reduce the selenodiazole ring back to the diamino benzene (**2-49**). [13]



Scheme 2.10 Nitration of diamino benzene

With compound **2-48** in hand it was possible to synthesize two of the benzimidazole derivatives **2-49** and **2-50** (**Scheme 2.11.A**). These products were then coupled with 3–chloro–5–fluoroaniline (**2-34**) to furnish the desired products. A third derivative in this series was produced through coupling with compound **2-51**, which was provided by a collaborator (**Scheme**

2.11.B). In addition to these compounds a forth member of this series replaced the benzoxadiazole ring with a benzene ring (*synthesized by Leila Bayeh, UTSW*) (**Scheme 2.11.C**)



Scheme 2.11 Synthesis of a series of A-ring heteroaromatic analogs

2.4.2 Activity of heteroaromatic derivatives

Compounds **2-52** – **2-55** were evaluated for biological activity via the AlphaScreen (**Table 2.3**). These compounds showed greatly reduced activity over the benzoxadiazole. The stronger activity of the parent compound is likely due to the unique electronic distribution of the benzoxadiazole ring.

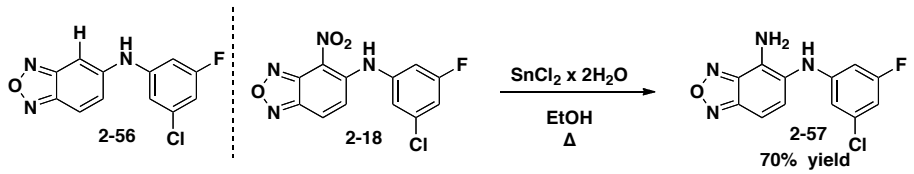
Table 2.3: Biological data for compounds **2-52** – **2-55**.

Compounds	AlphaScreen data (μM)	K_D (μM) ITC data	Compounds	AlphaScreen data (μM)	K_D (μM) ITC data
2-52	**ND**	**4.3**	**2-54**	**>5**	**–**
2-53	**>10**	**>20**	**2-55**	**>2**	**–**

2.5 SAR analysis: A–ring, nitro group

2.5.1 Synthesis of heteroaromatic ring structures

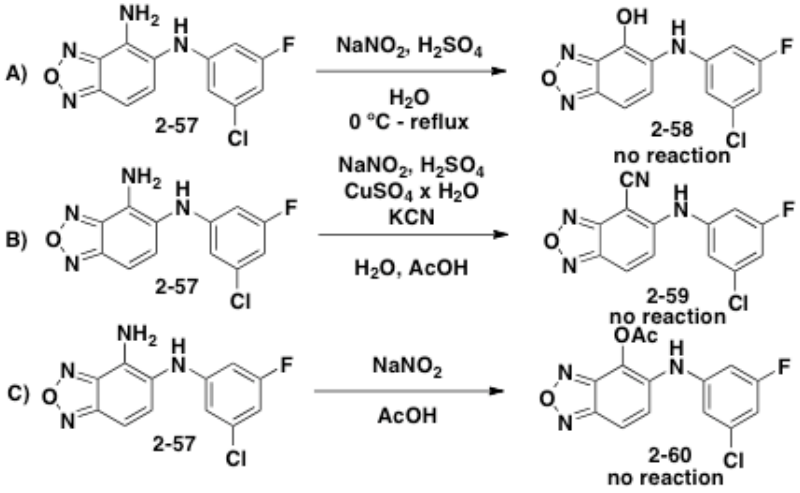
To evaluate the importance of the nitro–functionality on the A ring, a series of alterations were approached. Nitro groups on pharmaceutical compounds can present a potential liability as they can be metabolized in the body to give dangerous radicals. The nitro group was completely removed and replaced with a hydrogen atom (**2-56**, *compound synthesized by Leila Bayeh, UTSW*). The nitro group was also reduced to the corresponding amine using tin (II) chloride dihydrate to give compound **2-57** (**Scheme 2.12**). [14]



Scheme 2.12 Reduction and elimination of A-ring nitro group

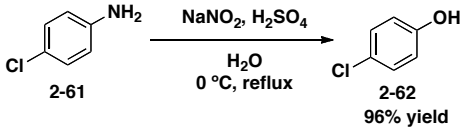
With the amine group as a handle, we looked at alterations we could make at this position. As illustrated in **Scheme 2.13**, our first attempt was to convert the amine to the corresponding hydroxyl group.[15] Using sodium nitrite and sulfuric acid in aqueous media this reaction produced only trace amounts of the desired product and the starting material was recovered almost quantitatively. Additionally, by using potassium cyanide we sought to obtain

the nitrile product but again the reaction gave almost exclusively starting material.[16] Finally, to convert the amine to the corresponding acetate the starting material was reacted with sodium nitrite and acetic acid and again the reaction was unsuccessful.[15]



Scheme 2.13 Efforts at conversion of A–ring amino group

To test the methodology, we attempted to convert 4–chloroaniline into the corresponding phenol using sodium nitrite and sulfuric acid. In this reaction, complete conversion to the desired product was observed within a few hours (**Scheme 2.14**). Again we hypothesized that the unique electronics of the benzoxadiazole ring caused the amine group to behave differently than a traditional aniline group.



Scheme 2.14 Aniline conversion to phenol

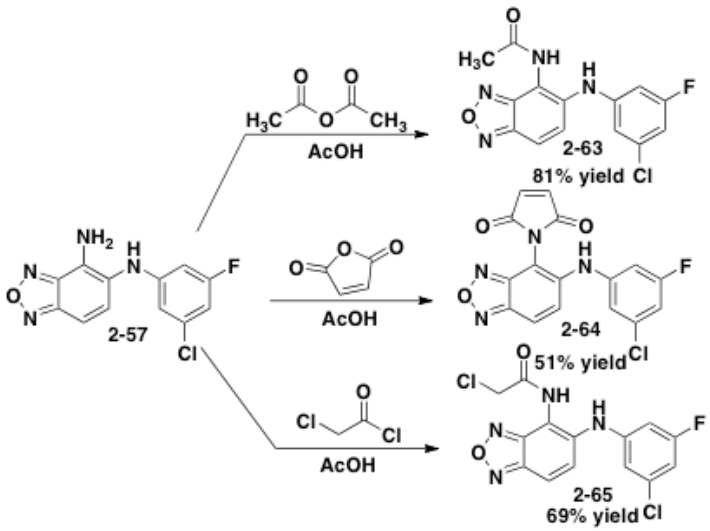
2.5.2 Activity: amine derivative

Although attempts to further manipulate the A–ring amine group were unsuccessful, the biological activity of the amine (**2-58**) was evaluated. With the amine replacing the nitro group, a significant decrease in activity was observed. It should be noted, however, that the poor activity of this compound could be due in part to its reduced solubility in the assay when compared to the nitro derivative.

Conversion of the A–ring nitro group to an amine represented a significant change in functionality. Not only was a strongly electron withdrawing group replaced by an electron donating group, but the nitro group represented a hydrogen–bond acceptor while the amine offered the ability to be a hydrogen bond donor.

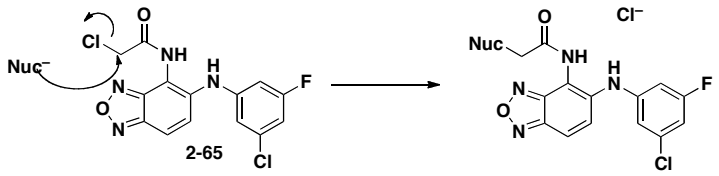
2.5.3 More derivatives of the A–ring nitro group

To evaluate analogs that more closely resembled the lead compound, a series of acetyl derivatives were synthesized (**Scheme 2.15**).



Scheme 2.15 Acylation of A-ring amino group

Not only did these compounds provide electron–withdrawing activity, they also represented potentially non–reversible binding ligands (**Scheme 2.16**).



Scheme 2.16 Irreversible binding with compound **2-65**.

The A–ring nitro group was also replaced with a sulfamido group (compounds **2-66** and **2-67**) as the sulfamido group was similar electronically to the nitro group (**Figure 2.3**). Finally, x–ray co–crystals showed additional space surrounding the A–ring. This provided the opportunity to expand on the parent structure and take advantage of additional interactions with the protein. Two additional compounds were synthesized that expanded on the original structure by adding a bromine atom at the 7–position (**2-68**) and a phenyl substituent also at the 7–position (**2-69**). (*Compounds **2-66** – **2-69** were synthesized by Leila Bayeh, UTSW*).

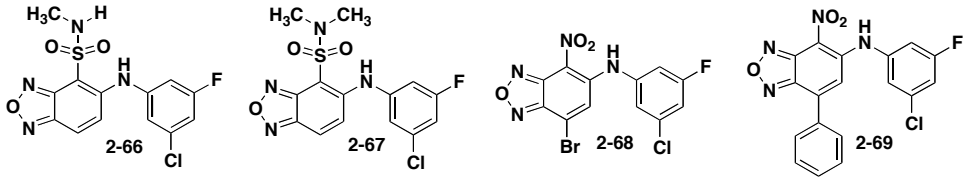


Figure 2.3 Structures of compounds **2-66** – **2-69**

The AlphaScreen for these compounds showed no activity at the concentration tested. Again these results support a unique electronic environment specific to the benzoxadiazole ring that provides the activity observed in the parent compound. Changes in the substitution to this ring generally result in a complete loss of activity.

Table 2.4: Biological data for compounds **2-56** – **2-69**

Compounds	AlphaScreen data (μM)	K_D (μM) ITC data	Compounds	AlphaScreen data (μM)	K_D (μM) ITC data
2-56	**NA**	**~3**	**2-66**	**NA**	**5**
2-57	**>30**	**~6**	**2-67**	**NA**	**5**
2-63	**NA**	**–**	**2-68**	**NA**	**–**
2-64	**NA**	**–**	**2-69**	**NA**	**–**
2-65	**NA**	**–**	**–**	**–**	**–**

2.6 Conclusion

This SAR study provided important insight into the structural requirements for an inhibitor of the HIF–2α heterodimerization. The results show that potent inhibitors bind within the internal cavity of the HIF–2α PAS–B domain, which results in a change in conformation of the protein. This in turn disrupts the binding interactions with the ARNT subunit. While various substituents can be tolerated around the B–ring of the molecule, the A–ring and linker are much less tolerant. This suggests an important electronic environment on this portion of the molecule that is necessary for binding.

These results are important in showing that a small molecule ligand can be used to modulate HIF activity and this approach can be used to study the protein–protein interaction of

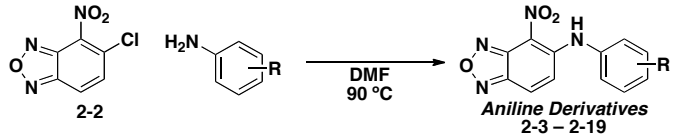
HIF. These small molecule inhibitors can also provide a way of studying HIF attenuation in vivo to better understand how this attenuation will influence cellular processes.

2.7 Experimental Section

2.6.1 Materials and Methods

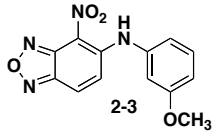
All reactions were performed under an atmosphere of nitrogen in flame-dried glassware and with dry solvents. Commercially obtained reagents were used as received. Reactions were monitored by thin-layer chromatography on silica gel plates (60 F254 pre-coated plates (0.25 mm). Flash chromatography was performed using silica gel (particle size 0.032-0.063 mm) purchased from SiliCycle. ^{1}H and ^{13}C NMR spectral data were recorded at 600 MHz or 400 MHz and 100 MHz respectively in $CDCl_3$, and chemical shifts were referenced to the corresponding solvent residual signal. Coupling constants are reported as three bond H-H coupling unless otherwise noted. Low-resolution LC/ESI-MS data were measured using ESI-MS with a reversed phase C_{18} column (Phenomenex Luna, 150 mm x 4.6 mm, 5 mm) at a flow rate of 0.7 mL/ min.

*2.6.2 Synthesis of Aniline Derivatives **2-3** – **2-19**:*



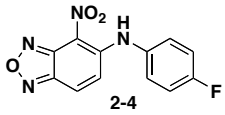
A flame-dried reaction vial was charged with benzoxadiazole **2-2** (50 mg, 0.25 mmol) and anhydrous DMF (1.5 mL). The mixture was treated with an aniline (0.25 mmol) and stirred at 90 °C for 3 h. After cooling to room temperature, the reaction was then diluted with ethyl acetate (5 mL) and washed with water (3 x 5 mL). The combined aqueous layers were extracted

with ethyl acetate (2 x 5 mL). The combined organic layers were washed with brine (10 mL), dried over $MgSO_4$, and concentrated under reduced pressure. The resulting powder was recrystallized from 30% ethyl acetate in hexanes to provide crystals of the desired aniline derivative.



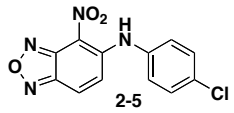
2-3

2-3: Following the typical procedure for the synthesis of aniline derivatives with *m*-anisidine, recrystallization provided yellow crystals of the desired product **2-3** (50.1 mg, 70% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 8.10 (d, *J* = 9.9 Hz, 1H), 7.50 (m, 2H), 7.14 (dd, *J* = 2.1 Hz, 2.1 Hz, 1H), 7.08 (m, 2H), 3.87 (s, 3H) ^{13}C NMR (100 MHz, Acetone-d_6) δ 160.9, 146.9, 145.1, 141.3, 130.6, 130.6, 130.1, 126.1, 123.8, 120.7, 118.5, 93.8, 55.1. LCMS (ESI) calc'd for $[C_{13}H_9N_4O_4]^-$([M-H]$^-$): *m/z* 285.1, found 285.0.



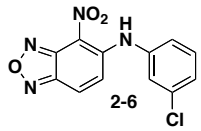
2-4

2-4: Following the typical procedure for the synthesis of aniline derivatives with 4-fluoroaniline, recrystallization provided orange crystals of the desired product **2-4** (60.9 mg, 88% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.73 (bs, 1H), 8.08 (d, *J* = 9.9 Hz, 1H), 7.60 (m, 2H), 7.44 (d, *J* = 9.9 Hz, 1H), 7.38 (m, 2H) ^{13}C NMR (100 MHz, Acetone-d_6) δ 163.0 (d, *J* = 252.0 Hz), 150.3, 147.8, 146.2, 130.0 (d, *J* = 8.9 Hz), 127.2, 127.1, 125.3, 117.6 (d, *J* = 23.0 Hz), 108.8. LCMS (ESI) calc'd for $[C_{12}H_6FN_4O_3]^-$([M-H]$^-$): *m/z* 273.0, found 273.0.



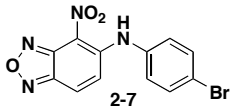
2-5

2-5: Following the typical procedure for the synthesis of aniline derivatives with 4-chloroaniline, recrystallization provided orange crystals of the desired product **2-5** (36.6 mg, 51% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.7 (bs, 1H), 8.10 (d, J = 9.9 Hz, 1H), 7.6 (m, 4H), 7.50 (d, J = 9.9 Hz, 1H) ^{13}C NMR (100 MHz, Acetone-d_6) δ 147.8, 146.2, 142.3, 134.3, 131.2, 131.2 130.9, 129.4, 127.2, 125.3. LCMS (ESI) calc'd for $[C_{12}H_6ClN_4O_3]^-$([M-H]): *m/z* 289.0, found 289.0.



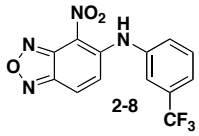
2-6

2-6: Following the typical procedure for the synthesis of aniline derivatives with 3-chloroaniline, recrystallization provided yellow crystals of the desired product **2-6** (54.4 mg, 75% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.7 (bs, 1H), 8.12 (d, J = 9.9 Hz, 1H), 7.65 (m, 1H), 7.61 (d, J = 9.9 Hz, 1H), 7.65 (m, 1H), 7.61 (d, J = 7.9 Hz, 1H), 7.52 (m, 3H) ^{13}C NMR (100 MHz, Acetone-d_6) δ 148.5, 146.9, 145.1, 138.4, 138.3, 134.7, 131.3, 128.1, 126.7, 126.3, 125.3, 124.3. LCMS (ESI) calc'd for $[C_{12}H_6ClN_4O_3]^-$([M-H]$^-$): *m/z* 289.0, found 289.0.



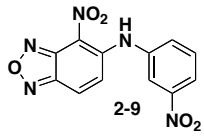
2-7

2-7: Following the typical procedure for the synthesis of aniline derivatives with 4-bromoaniline, recrystallization provided orange crystals of the desired product **2-7** (55.5 mg, 67% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.7 (bs, 1H), 8.10 (d, J = 9.9 Hz, 1H), 7.77 (d, J = 8.8 Hz, 2H), 7.52 (m, 3H) ^{13}C NMR (100 MHz, Acetone-d_6) δ 147.8, 146.0, 134.0, 133.0, 127.4, 127.4, 127.3, 125.3, 124.8, 122.1 LCMS (ESI) calc'd for $[C_{12}H_6BrN_4O_3]^-$([M-H]$^-$): *m/z* 333.0, found 333.0.



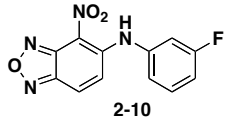
2-8

2-8: Following the typical procedure for the synthesis of aniline derivatives with 3-trifluoromethyl aniline, recrystallization provided yellow crystals of the desired product **2-8** (43.8 mg, 54% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.80 (bs, 1H), 8.10 (d, J = 9.8 Hz, 1H), 7.95 (s, 1H), 7.87 (m, 1H), 7.84 (m, 2H), 7.53 (d, J = 9.8 Hz, 1H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 148.8, 146.8, 145.1, 137.9, 131.6 (q, J = 32.5 Hz), 131.0, 130.7, 126.1, 124.7 (q, J = 3.8 Hz), 124.4, 123.8 (q, J = 271 Hz), 123.7 (q, J = 3.9 Hz), 120.0 LCMS (ESI) calc'd for $[C_{13}H_6F_3N_4O_3]^-$([M-H]$^-$): *m/z* 323.0, found 323.0.



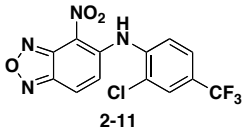
2-9

2-9: Following the typical procedure for the synthesis of aniline derivatives with 3-nitroaniline, recrystallization provided red crystals of the desired product **2-9** (45.9 mg, 61% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.81 (bs, 1H), 8.47 (dd, *J* = 2.1 Hz, 2.1 Hz, 1H), 8.34 (d, *J* = 8.2 Hz, 1H), 8.13 (d, *J* = 9.9 Hz, 1H), 8.04 (d, *J* = 9.0 Hz, 1H), 7.90 (dd, *J* = 8.1 Hz, 8.1 Hz, 1H), 7.60 (d, *J* = 9.8 Hz, 1H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 146.9, 145.7, 138.3, 133.0, 131.1, 128.3, 126.2, 124.8, 124.5, 123.3, 122.6, 121.8. LCMS (ESI) calc'd for $[C_{12}H_6N_5O_5]^-$([M-H]$^-$): *m/z* 300.0, found 300.0.



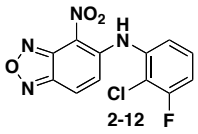
2-10

2-10: Following the typical procedure for the synthesis of aniline derivatives with 3-fluoroaniline, recrystallization provided yellow crystals of the desired product **2-10** (25.3 mg, 37% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.7 (bs, 1H), 8.12 (d, *J* = 9.8 Hz, 1H), 7.65 (dd, *J* = 7.2 Hz, 8.0 Hz, 1H), 7.55 (d, 1H, *J* = 9.9 Hz), 7.41 (d, *J* = 7.2 Hz, 2H), 7.29 (ddd, *J* = 2.2 Hz, 8.7 Hz, 8.7 Hz, 1H), ^{13}C NMR (100 MHz, Acetone-d_6) δ 164.1 (d, *J* = 245.0 Hz), 149.7, 149.7, 147.8, 146.1, 132.5 (d, *J* = 9.3 Hz), 131.6, 127.2, 125.3, 123.6 (d, *J* = 3.1 Hz), 115.9 (d, 21.1 Hz), 114.9 (d, 23.8 Hz). LCMS (ESI) calc'd for $[C_{12}H_6FN_4O_3]^-$([M-H]$^-$): *m/z* 273.0, found 273.0.



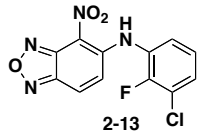
2-11

2-11: Following the typical procedure for the synthesis of aniline derivatives with 2-chloro-4-trifluoromethyl aniline, purification by flash chromatography on silica gel (20% ethyl acetate in hexanes) provided the desired product **2-11** as an orange powder (67.1 mg, 75% yield) ^{1}H NMR (400 MHz, Acetone-d_6) δ 9.44 (bs, 1H), 8.68 (d, J = 8.8 Hz, 1H), 8.59 (s, 1H), 7.94 (dd, J = 6.8 Hz, 8.5 Hz, 1H), 7.84 (s, 1H), 7.71 (d, J = 8.7 Hz, 1H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 146.9, 145.0, 138.2, 131.8, 127.7 (q, J = 3.9 Hz), 126.3 (q, J = 3.5 Hz), 126.2 (q, J = 278.0 Hz), 126.0, 125.5 (q, J = 5.1 Hz), 124.8 (q, J = 3.7 Hz), 124.8, 121.6, 121.5. LCMS (ESI) calc'd for $[C_{13}H_5ClF_3N_4O_3]^-$([M-H]$^-$): *m/z* 357.0, found 357.0.



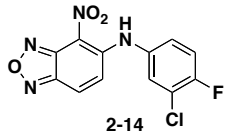
2-12

2-12: Following the typical procedure for the synthesis of aniline derivatives with 2-chloro-3-fluoroaniline, recrystallization provided yellow crystals of the desired product **2-12** (50.4 mg, 66% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.6 (bs, 1H), 8.15 (d, J = 9.9 Hz, 1H), 7.60 (m, 2H), 7.50 (ddd, J = 1.6 Hz, 8.8 Hz, 8.8 Hz, 1H), 7.39 (dd, J = 2.7 Hz, 9.8 Hz, 1H) ^{13}C NMR (100 MHz, Acetone-d_6) δ 158.9 (d, J = 248.0 Hz), 148.4, 146.9, 145.0, 136.2, 136.1, 129.0 (d, J = 9.1 Hz), 126.8, 126.0, 124.7 (d, J = 4.2 Hz), 118.8 (d, J = 18.5 Hz), 116.4 (d, J = 21.1 Hz) LCMS (ESI) calc'd for $[C_{12}H_5ClFN_4O_3]^-$([M-H]$^-$): *m/z* 307.0, found 307.0.



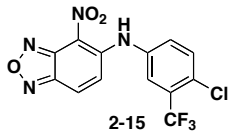
2-13

2-13: Following the typical procedure for the synthesis of aniline derivatives with 3-chloro-2-fluoroaniline, recrystallization provided yellow crystals of the desired product **2-13** (51.3 mg, 67% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.5 (bs, 1H), 8.15 (d, J = 9.9 Hz, 1H), 7.67 (m, 2H), 7.46 (m, 2H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 153.5 (d, J = 249.0 Hz), 147.9, 146.0, 135.1, 131.5, 129.0, 127.0, 127.0, 126.7 (d, J = 5.2 Hz), 125.7, 122.7 (d, J = 16.3 Hz), 96.0. LCMS (ESI) calc'd for $[C_{12}H_5ClFN_4O_3]^-$([M-H]$^-$): *m/z* 307.0, found 307.0.



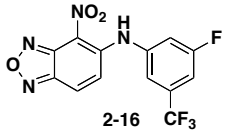
2-14

2-14: Following the typical procedure for the synthesis of aniline derivatives with 3-chloro-4-fluoroaniline, recrystallization provided yellow crystals of the desired product **2-14** (42.3 mg, 55% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.68 (bs, 1H), 8.09 (d, J = 9.9 Hz, 1H), 7.96 (s, 1H), 7.80 (dd, J = 2.3 Hz, 6.7 Hz, 1H), 7.60 (m, 1H), 7.55 (m, 2H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 162.5, 158.0 (d, J = 250.0 Hz), 149.7, 147.4, 145.7, 134.7, 130.0, 128.4 (d, J = 7.7 Hz), 126.9, 125.0, 121.9 (d, J = 19.0 Hz), 118.4 (d, J = 18.4 Hz). LCMS (ESI) calc'd for $[C_{12}H_5ClFN_4O_3]^-$([M-H]$^-$): *m/z* 307.0, found 307.0.



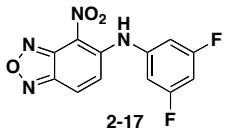
2-15

2-15: Following the typical procedure for the synthesis of aniline derivatives with 4-chloro-3-trifluoromethylaniline, recrystallization provided orange crystals of the desired product **2-15** (33.6 mg, 38% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.73 (bs, 1H), 8.13 (d, J = 9.9 Hz, 1H), 8.08 (s, 1H), 7.89 (d, J = 1.4 Hz, 1H), 7.62 (d, J = 9.9 Hz, 1H), ^{13}C NMR (100 MHz, Acetone-d_6) δ 147.9, 146.0, 138.7, 137.5, 134.0, 133.0, 131.6 (q, J = 1.9 Hz), 129.8 (q, J = 31.6 Hz), 127.4 (q, J = 5.2 Hz), 122.3, 125.3, 123.6 (q, J = 247.2 Hz), 120.9. LCMS (ESI) calc'd for $[C_{13}H_5ClF_3N_4O_3]^-$([M-H]$^-$): *m/z* 357.0, found 357.0.



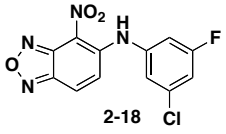
2-16

2-16: Following the typical procedure for the synthesis of aniline derivatives with 3-fluoro-5-trifluoromethylaniline, recrystallization provided orange crystals of the desired product **2-16** (13.4mg, 26% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.67 (bs, 1H), 8.16 (d, J = 9.9 Hz, 1H), 7.84 (s, 1H), 7.77 (ddd, J = 2.2 Hz, 2.2 Hz, 9.3 Hz, 1H), 7.66 (d, J = 9.8 Hz, 1H), 7.66 (d, J = 8.5 Hz, 1H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 163.0 (d, J = 260 Hz), 148.2, 146.9, 145.0, 140.0 (d, J = 11.6 Hz), 138.1 (q, J = 268.8 Hz), 133.0, 126.4, 124.5, 121.8 (d J = 3.2 Hz), 119.9 (dq, J = 3.4, 3.7 Hz), 118.0 (d, J = 23.8 Hz), 112.1 (dq, J = 3.7 Hz, 30.0 Hz). LCMS (ESI) calc'd for $[C_{13}H_5F_4N_4O_3]^-$([M-H]$^-$): *m/z* 341.0, found 341.0.



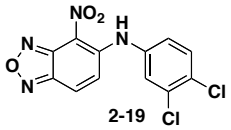
2-17

2-17: Following the typical procedure for the synthesis of aniline derivatives with 3,5-difluoroaniline, recrystallization provided orange crystals of the desired product **2-17** (52.6 mg, 72% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.65 (bs, 1H), 8.15 (d, J = 9.8 Hz, 1H), 7.66 (dd, J = 3.1 Hz, 9.9 Hz, 1H), 7.32 (d, J = 5.8 Hz, 2H), 7.18 (ddd, J = 2.4 Hz, 2.4 Hz, 9.9 Hz, 1H), ^{13}C NMR (100 MHz, Acetone-d_6) δ 163.5 (d, J = 243.9 Hz), 163.3 (d, J = 246.7 Hz), 148.3, 146.9, 145.0, 139.7 (dd, J = 12.7 Hz, 13.0 Hz), 126.4, 124.5, 110.3 (d, J = 27.8 Hz), 110.3 (d, J = 11.4 Hz), 103.3 (dd, J = 18.0 Hz, 33.0 Hz), 94.9. LCMS (ESI) calc'd for $[C_{12}H_5F_2N_4O_3]^-$([M-H]$^-$): m/z 291.0, found 291.0.



2-18

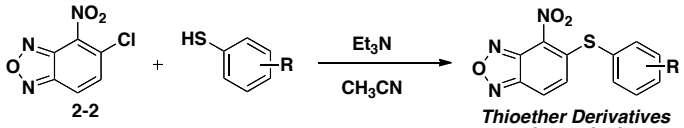
2-18: Following the typical procedure for the synthesis of aniline derivatives with 3-chloro-5-fluoroaniline, recrystallization provided yellow crystals of the desired product **2-18** (436.6 mg, 81% yield): ^{1}H NMR (400 MHz, $CDCl_3$) δ 11.52 (bs, 1H), 7.97 (d, J = 9.8 Hz, 1H), 7.35 (d, J = 9.8 Hz, 1H), 7.21 (ddd, J = 1.9 Hz, 8.1 Hz, 1H), 7.16 (bs, 1H), 6.99 (ddd, J = 1.9 Hz, 8.1, 1H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 163.2 (d, J = 253.0 Hz), 147.6, 146.6, 144.7, 141.1, 138.3, 137.2 (d, J = 10.7 Hz), 125.5, 124.5, 122.4 (d, J = 3.6 Hz), 116.8 (d, J = 24.4 Hz), 112.0 (d, J = 26.5 Hz). LCMS (ESI) calc'd for $[C_{12}H_5ClFN_4O_3]^-$([M-H]$^-$): m/z 307.0, found 307.0.



2-19

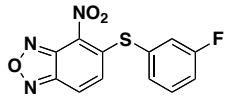
2-19: Following the typical procedure for the synthesis of aniline derivatives with 3,4-dichloroaniline, recrystallization provided yellow crystals of the desired product **2-19** (51.2 mg, 63% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 11.7 (bs, 1H), 8.13 (d, J = 9.9 Hz, 1H), 7.84 (d, J = 2.4 Hz, 1H), 7.79 (d, J = 8.6 Hz, 1H), 7.59 (m, 2H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 149.6, 147.9, 146.1, 138.0, 133.8, 132.6, 132.3, 129.8, 127.9, 127.3, 125.3, 107.9. LCMS (ESI) calc'd for $[C_{12}H_5Cl_2N_4O_3]^-$([M-H]$^-$): *m/z* 323.0, found 323.0.

*2.6.3 Synthesis of Thiophenol Derivatives **2-30** – **2-32**:*



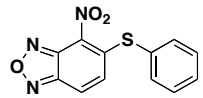
2-2 *Thioether Derivatives 2-30 – 2-32*

A flame-dried reaction vial was charged with benzoxadiazole **2** (50 mg, 0.25 mmol) and anhydrous CH_3CN (1.5 mL). The mixture was treated with a thiophenol (0.25 mmol) and triethylamine (25 mg, 0.25 mmol). The reaction was stirred at ambient temperature and monitored by LCMS. At the completion of the reaction, the solvent was removed under a stream of nitrogen gas. Purifciation by flash chromatography on silica gel (1:1 hexane:dichloromethane) provided the desired thiophenol derivative.



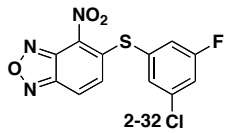
2-30

2-30: Following the typical procedure for the synthesis of thiophenol derivatives with 3-fluorothiobenzene, purification by flash chromatography on silica gel provided the desired product **2-30** as a yellow solid (15.8 mg, 20% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 8.15 (d, *J* = 9.6 Hz, 1H), 7.71 (m, 1H), 7.62 (m, 2H), 7.48 (m, 1H), 7.29 (d, *J* = 9.6 Hz, 1H) ^{13}C NMR (100 MHz, Acetone-d_6) δ 168.3 (d, *J* = 256.0 Hz), 154.6 (d, *J* = 22.5 Hz), 149.6, 137.6 (d, *J* = 8.4 Hz), 137.0, 136.8 (d, *J* = 3.2 Hz), 136.7, 128.2, 127.3 (d, *J* = 22.6 Hz), 126.6, 126.6, 123.5 (d, *J* = 21.0 Hz). LCMS (ESI) calc'd for $[C_{12}H_5FN_3O_3]^-$([M-H]$^-$): *m/z* 290.0, found 290.0.



2-31

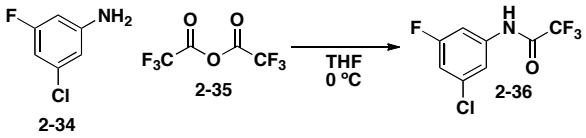
2-31: Following the typical procedure for the synthesis of thiophenol derivatives with thiobenzene, purification by flash chromatography on silica gel provided the desired product **2-31** as a yellow solid (42.5 mg, 58% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 8.13 (d, J = 9.6 Hz, 1H), 7.78 (dd, J = 1.4 Hz, 7.8 Hz, 2H), 6.67 (m, 3H), 7.21 (d, J = 9.6 Hz, 1H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 150.3, 149.2, 144.4, 135.5, 131.5, 131.3, 130.7, 129.3, 121.2, 98.8 LCMS (ESI) calc'd for $[C_{12}H_6N_3O_3S]^-$([M-H]$^-$): m/z 272.0, found 272.0.



2-32

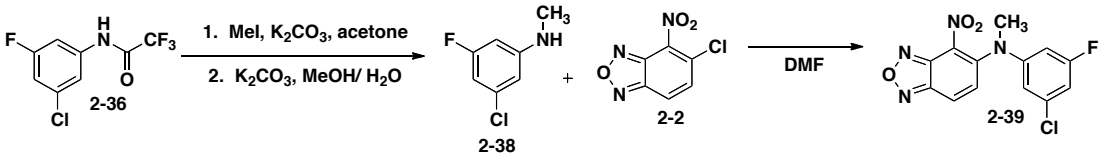
2-32: Following the typical procedure for the synthesis of thiophenol derivatives with 3-chloro-5-fluorothiobenzene, purification by flash chromatography on silica gel provided the desired product **2-32** as a yellow solid (123.6 mg, 78% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 7.84 (d, *J* = 9.7 Hz, 1H), 7.21 (d, *J* = 9.8 Hz, 1H), 6.96 (m, 2H), 6.80 (d, *J* = 9.4 Hz, 1H), 3.37 (s, 3H) ^{13}C NMR (100 MHz, Acetone-d_6) δ 163.0 (d, *J* = 252 Hz), 149.3, 148.6, 144.3, 136.4 (d, *J* = 10.8 Hz), 133.3 (d, *J* = 8.9 Hz), 132.1, 131.2 (d, *J* = 3.5 Hz), 122.3, 121.7, 121.0 (*J* = 22.7 Hz), 118.8 (*J* = 24.9 Hz) LCMS (ESI) calc'd for $[C_{12}H_6ClFN_3O_3S]^+$([M+H]$^+$): m/z 326.0, found 326.0.

*2.6.4 Methylation of the amine linker **2-39**.*



2-36: A solution of 3-chloro-5-fluoroaniine (270 mg, 1.8 mmol) in THF (10 mL) was cooled to 0 °C and treated with trifluoroacetic anhydride (770 mg, 3.68 mmol). After 1 h, the solvent was removed under reduced pressure, and and the crude oil was dissolved in ethyl acetate. The organic layer was washed with saturated aqueous $NaHCO_3$ (3x), dried over $MgSO_4$ and concentrated under reduced pressure to provide trifluroacetamide **2-36** as a white solid (710 mg, 82%): ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.92 (bs, 1H), 7.39 (d, *J* = 12.7 Hz, 1H), 7.13 (s, 1H), 6.86 (ddd, *J* = 1.7 Hz, 1.7 Hz, 8.2 Hz, 1H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 162.7 (d, *J* = 248.6

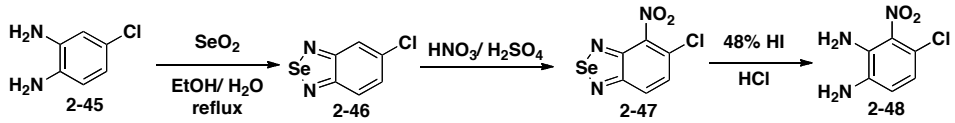
Hz), 155.0 (q, *J* = 37.8 Hz), 136.9 (d, *J* = 11.9 Hz), 135.8 (d, *J* = 11.9 Hz), 116.8 (q, *J* = 285.0 Hz), 116.3 (d, *J* = 3.5 Hz), 114.1 (d, *J* = 24.7 Hz), 106.5 (d, *J* = 26.5 Hz) LCMS (ESI) calc'd for $[C_8H_5ClF_4NO]^+$($[M+H]^+$): *m/z* 242.0, found 242.0.



2-39: A solution trifluroacetamide **2-36** in anhydrous acetone was sequentially treated with K_2CO_3 and methyl iodide. The reaction mixture was heated to reflux for 2 h and then filtered by vacuum filtration. The the filtrate was concentrated under reduced pressure and dissolved in CH_2Cl_2. The solution was washed with H_2O and brine, dried over $MgSO_4$ and concentrated under reduced pressure to yield a yellow solid (550 mg). The crude material was dissolved in 2:1 MeOH:H_2O (3 mL), and treated with potassium carbonate (430 mg, 3.1 mmol). The reaction was stirred for 12 h and then diluted with CH_2Cl_2. The mixture was washed with H_2O and brine, dried over $MgSO_4$ and concentrated under reduced pressure to provide methyl aniline **2-38** as a clear oil. This oil was dissolved in DMF (1.5 mL) and treated with benzoxadiazole **2** (79 mg, 0.39 mmol). After 4 h, the reaction mixture was diluted with ethyl acetate and washed repeatedly with H_2O. The organic layer was dried over $MgSO_4$ and concentrated under reduced pressure. The crude material was purified by flash chromatography on silica gel (20% ethyl acetate in hexanes) to provide aniline **2-39** as an orange solid (129 mg , 85 % yield): ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.84 (d, *J* = 9.7 Hz, 1H), 7.21 (d, *J* = 9.8 Hz, 1H), 6.96 (m, 2H), 6.80 (d, *J* = 9.4 Hz, 1H), 3.37 (s, 3H) ^{13}C NMR (100 MHz, $CDCl_3$) δ 163.5 (d, *J* =

250.7 Hz), 147.1 (d, *J* = 11.3 Hz), 146.9, 145.3, 137.0 (d, *J* = 12.3 Hz), 131.5, 125.5, 124.5, 121.3 (d, *J* = 11.3 Hz), 119.1 (d, *J* = 3.3 Hz), 114.2 (d, *J* = 22.2 Hz), 108.9 (d, *J* = 24.3 Hz), 42.6. LCMS (ESI) calc'd for $[C_{13}H_9ClFN_4O_3]^-$([M-H]$^-$): *m/z* 323.0, found 323.0.

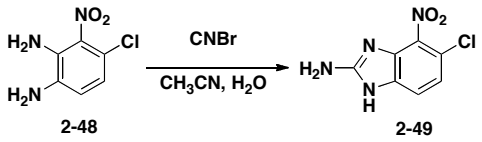
2.6.5 Alterations to the heterocycle ***2-52*** *–* ***2-54****:*



2-48 was prepared following a known literature procedure[1]: A solution of 3-chloroorthophenylene diamine **2-45** (252 mg, 1.77 mmol) in ethanol (2.0 mL) was heated to reflux and treated dropwise with a solution of selenium dioxide (216 mg, 1.94 mmol) in water (1 mL). The reaction was monitored by TLC. After 30 min, the mixture was cooled to ambient temperature and filtered via vacuum filtration to give the desired selenadiazole as a dark brown solid (253 mg, 66%). The crude product was carried forward to the next step without further purification: ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.99 (s, 1H), 7.86 (d, *J* = 6.2 Hz, 1H), 7.55 (d, *J* = 5.9 Hz, 1H). LCMS (ESI) calc'd for $[C_6H_4ClN_2Se]^+$([M+H]$^+$): *m/z* 218.9, found 218.9.

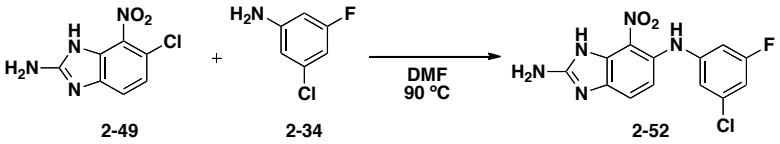
The selenadiazole (157 mg, 0.726 mmol) from the previous step was dissolved in conc. H_2SO_4 (2.4 mL) and cooled in an ice water bath. The dark green solution was treated dropwise with conc. HNO_3 (0.16 mL) and turned dark red in color. After 50 min, the reaction mixture was poured onto ice and filtered via vacuum filtration to yield the nitrated selenadiazole product as a light brown powder (145 mg, 77% yield). The crude product was carried forward to the next step without further purification: ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.13 (d, *J* = 7.6 Hz, 1H), 7.82 (d, *J* = 7.6 Hz, 1H). LCMS (ESI) calc'd for $[C_6H_3ClN_3O_2Se]^+$([M+H]$^+$): *m/z* 263.9, found 263.5.

The nitroselenadiazole (83.54 mg, 0.321 mmol) from the previous step was dissolved in conc. HCl (0.78 mL) and cooled in an ice bath. The reaction mixture was treated with a 48% HI solution (0.26 mL) followed by 50% NaOH to get to a pH of 8. The product was extracted with ethyl acetate. The organic layer was washed with brine and concentrated under reduced pressure to provide **2-48** as a red powder (45.7 mg, 76%): ^{1}H NMR (400 MHz, Acetone-d_6) δ 6.78 (d, J = 8.2 Hz, 1H), 6.68 (d, J = 8.2 Hz, 1H), 5.33 (bs, 2H), 5.18 (bs, 2H). LCMS (ESI) calc'd for $[C_6H_7ClN_3O_2]^+$($[M+H]^+$): *m/z* 188.0, found 188.0.

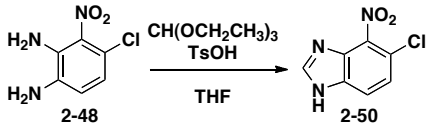


2-48 2-49

2-49 was prepared following a known literature procedure[2]: **2-48** (75 mg, 0.40 mmol) was dissolved in a 5:1 mixture of CH_3CN and water (1.2 mL) and cooled to 0 ºC. The dark red solution was treated with cyanogen bromide (47 mg, 0.44 mmol), and the reaction was monitored by LCMS. At the completion of the reaction, the solvent was removed under a stream of nitrogen gas. Conc. ammonium hydroxide was added, which resulted in a dark red precipitate formation that was filtered via vacuum filtration. The precipitate was washed repeatedly with cold water and dried under vacuum. Purified via flash chromatography on silica gel (15% ethyl acetate in hexanes) provided the desired product **2-49** as a red powder (35.0 mg, 41 %): ^{1}H NMR ($CDCl_3$, 400 MHz) δ 6.82 (bs, 2H), 6.78 (d, J = 8.2 Hz, 1H), 6.68 (d, J = 8.2 Hz, 1H), 5.33 (bs, 1H). LCMS (ESI) calc'd for $[C_7H_4ClN_4O_2]^-$($[M-H]^-$): *m/z* 211.0, found 211.0.

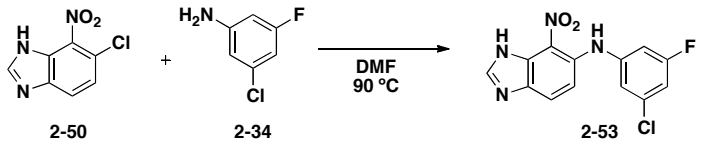


2-52: Following the typical procedure for the synthesis of aniline derivatives with chloride **2-49**, purification by flash chromatography on silica gel (20% ethyl acetate in hexanes) provided the desired product **2-52** as a yellow solid (19.7 mg, 65%): ^{1}H NMR (400 MHz, Acetone-d_6) δ 7.32 (d, *J* = 6.0 Hz, 1H), 7.28 (bs, 1H), 6.57 (s, 1H), 6.54 (ddd, *J* = 1.2 Hz, 1.3 Hz, 6.7 Hz, 1H), 6.40 (ddd, *J* = 1.3 Hz, 1.3 Hz, 6.7 Hz, 1H), 5.77 (s, 1H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 164.8 (d, *J* = 249.3 Hz), 150.0, 149.8 (d, *J* = 12.1 Hz), 136.1, 136.0, 134.0 (d, *J* = 13.6 Hz), 131.7, 118.8, 118.4 (d, *J* = 1.9 Hz), 110.9, 105.9 (d, *J* = 26.0 Hz), 103.3, 100.3 (d, *J* = 25.6 Hz). LCMS (ESI) calc'd for $[C_{13}H_7ClFN_5O_2]^-$([M-H]$^-$): *m/z* 320.0, found 320.1.

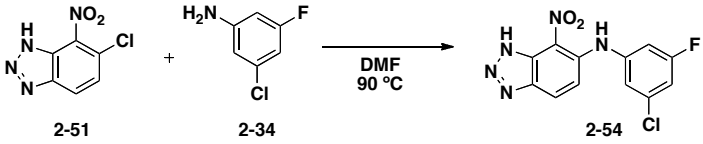


2-50 was prepared following a known literature procedure[2]: A solution of **2-48** (75 mg, 0.40 mmol) in THF (1.5 mL) was treated sequentially with triethyl orthoformate (178 mg, 1.2 mmol) and *p*-toluenesulfonic acid (7.6 mg, 200 μL, 0.04 mmol). The reaction mixture was stirred at 50 °C, the reaction was monitored by LCMS. After 2 h, the solvent was removed under a stream of nitrogen gas, and the resulting residue was partitioned between ethyl acetate and water. The organic layer was washed with saturated aqueous Na_2CO_3, dried over $MgSO_4$, and concentrated under reduced pressure to provide **2-50** as a light brown solid (57 mg, 72%): ^{1}H

NMR (400 MHz, Acetone-d_6) δ 8.42 (s, 1H), 8.03 (d, J = 8.2 Hz, 1H), 7.87 (d, J = 8.2 Hz, 1H), 7.48 (bs, 1H). LCMS (ESI) calc'd for $[C_7H_5ClN_3O_2]^+$($[M+H]^+$): m/z 198.0, found 198.0.



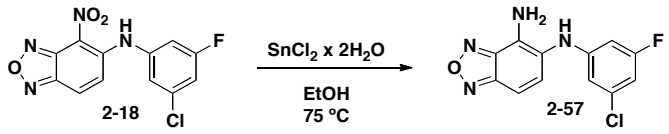
2-53: Following the typical procedure for the synthesis of aniline derivatives with chloride **2-50**, purification by flash chromatography on silica gel (15% ethyl acetate in hexanes) provided the desired product **2-52** as a yellow solid (9.4 mg, 31%): ^{1}H NMR (400 MHz, Acetone-d_6) δ 8.37 (d, J = 8.8 Hz, 1H), 8.01 (s, 1H), 7.72 (d, J = 8.8 Hz, 1H), 6.51 (dd, J = 1.8 Hz, 1.8 Hz, 1H), 6.34 (m, 2H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 164.8 (d, J = 241.6 Hz), 152.3 (d, J = 12.8 Hz), 142.6, 139.5, 135.7, 135.6 (d, J = 14.0 Hz), 129.9, 128.0, 127.1, 125.2, 118.6, 103.9 (d, J = 25.8 Hz), 100.0 (d, J = 24.5 Hz). LCMS (ESI) calc'd for $[C_{13}H_7ClFN_4O_2]^-$ ($[M-H]^-$): m/z 305.0, found 305.0.



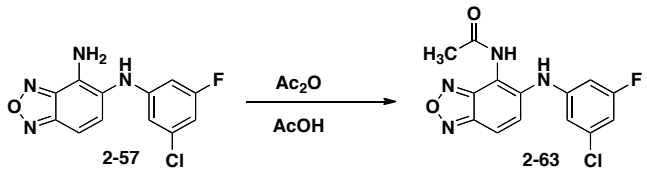
2-54: Following the typical procedure for the synthesis of aniline derivatives with chloride **2-51**, purification by flash chromatography on silica gel (10% ethyl acetate in hexanes) provided the desired product **2-54** as a yellow solid (9.4 mg, 16 %): ^{1}H NMR (400 MHz, Acetone-d_6) δ 7.96 (d, J = 1.2 Hz, 2H), 7.81 (d, J = 5.8 Hz, 1H), 7.62 (d, J = 3.6 Hz, 1H), 7.37

(dd, *J* = 1.2 Hz, 5.7 Hz, 1H,), 7.30 (d, *J* = 3.6 Hz, 1H), 6.4 (s, 1H), 6.23 (m, 1H), 5.14 (bs, 1H) ^{13}C NMR (100 MHz, Acetone-d_6) δ 164.8 (d, *J* = 242.0 Hz), 152.4 (d, *J* = 12.8 Hz), 145.4, 135.6 (d, *J* = 13.9 Hz), 135.2, 126.1, 124.0, 123.3, 110.1, 103.9 (d, *J* = 25.8 Hz), 99.9 (d, *J* = 24.6 Hz), 95.9. LCMS (ESI) calc'd for $[C_{12}H_6ClFN_5O_2]^-$([M-H]$^-$): *m/z* 306.0, found 306.0.

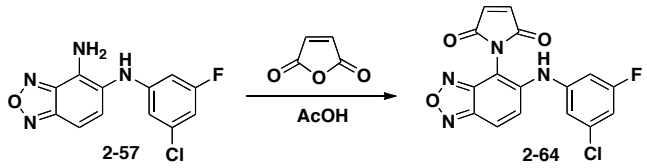
2.6.6 Alterations to nitro functionality:



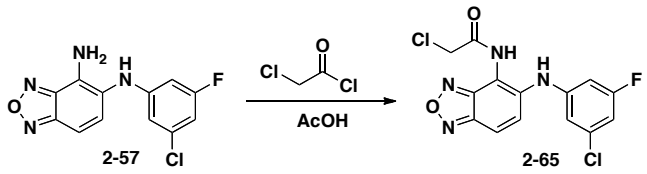
2-57: A solution of aniline **2-18** (29 mg, 0.094 mmol) in anhydrous EtOH (2 mL) was treated with $SnCl_2$•$2H_2O$ (63 mg, 0.28 mmol, 3 equiv). The reaction was heated to reflux for 4 h and then quenched with saturated aqueous $NaHCO_3$. The mixture was diluted with ethyl acetate passed through a pad of celite. The filtrate was washed with brine, dried over $MgSO_4$ and concentrated under reduced pressure. The crude oil was purified by flash chromatography on silica gel (20% ethyl acetate in hexanes) to provide 4-aminobenzoxadiazole **2-57** as an orange powder (18 mg, 70%): ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.17 (dd, *J* = 6.0 Hz, 12.0 Hz, 2H), 6.57 (d, *J* = 6.0 Hz, 1H), 6.44 (m, 2H), 6.25 (m, 1H), 5.26 (bs, 1H), 4.66 (bs, 2H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 164.0 (d, *J* = 246.0 Hz), 149.1, 147.0 (d, *J* = 11.3 Hz), 145.4, 136.2 (*J* = 12.9 Hz), 134.2, 130.3, 118.4, 110.6, 107.6 (d, *J* = 25.2 Hz), 107.5 (d, *J* = 27.9 Hz), 105.0. LCMS (ESI) calc'd for $[C_{12}H_7ClFN_4O]^-$([M-H]$^-$): *m/z* 277.0, found 277.1.



2-63: A solution of 4-aminobenzoxadiazole **2-57** (15.0 mg, 0.054 mmol) in acetic acid (1 mL) was treated with acetic anhydride (5.5 mg, 0.054 mmol). The reaction was stirred at ambient temperature for 19 h and quenched with ice. The desired acetamide **2-63** was collected by vacuum filtration as a green precipitate (14.0 mg, 81% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 9.54 (bs, 1H), 8.00 (bs, 1H), 7.83 (d, *J* = 6.4 Hz, 1H), 7.66 (d, *J* = 6.4 Hz, 1H), 6.85 (s, 1H), 6.81 (d, *J* = 5.7 Hz, 1H), 6.74 (d, *J* = 7.1 Hz, 1H) 2.25 (s, 3H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 169.5, 163.4 (d, *J* = 240.0 Hz), 147.9, 145.1 (d, *J* = 11.0 Hz), 135.3, 135.2, 130.2, 113.6, 113.9, 108.4, 108.2, 103.6, 103.3, 22.3. (LCMS (ESI) calc'd for $[C_{14}H_9ClFN_4O_2]^-$([M-H]$^-$): *m/z* 319.0, found 319.0.



2-64: Following the procedure for the synthesis of 4-amidobenzoxadiazole **2-57** with maleic anhydride, purification by vacuum filtration provided the acetamide **2-64** as a green precipitate (13.4 mg, 69% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 8.42 (bs, 1H), 7.89 (d, *J* = 6.4 Hz, 1H), 7.72 (d, *J* = 6.4 Hz, 1H), 6.93 (s, 1H), 6.81 (m, 3H), 6.39 (d, *J* = 8.4 Hz, 1H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 166.1, 162.8 (d, *J* = 243 Hz), 149.2, 148.8, 145.5, 145.4, 139.3, 136.2 (d, *J* = 13 Hz), 134.5, 133.3, 116.6, 115.7 (d, *J* = 5.8 Hz), 110.0 (d, *J* = 25.6 Hz), 105.4 (d, *J* = 25.3 Hz). LCMS (ESI) calc'd for $[C_{16}H_7ClFN_4O_3]^-$([M-H]$^-$): *m/z* 357.0, found 357.0.



2-65: Following the procedure for the synthesis of 4-amidobenzoxadiazole **2-57** with chloroacetylchloride, purification by reverse phase HPLC on a C-18 column (10% acetonitrile in water to 90% acetonitrile in water gradient eluent) provided the acetamide **2-65** as a green powder. (19.6 mg, 51% yield): ^{1}H NMR (400 MHz, Acetone-d_6) δ 8.08 (bs, 1H), 7.88 (d, *J* = 9.6 Hz, 1H), 7.68 (d, *J* = 9.6 Hz, 1H), 6.90 (s, 1H), 6.86 (d, *J* = 8.6 Hz, 1H), 6.78 (d, *J* = 10.6 Hz, 1H), 4.39 (s, 2H). ^{13}C NMR (100 MHz, Acetone-d_6) δ 165.8, 164.4 (d, *J* = 244.0 Hz), 147.9, 144.7 (d, *J* = 11.9 Hz), 143.9, 135.2 (d, *J* = 10.2 Hz), 129.9, 115.5, 114.7, 114.5, 108.9 (d, *J* = 24.5 Hz), 104.3 (d, *J* = 25.0 Hz), 95.0, 42.5. LCMS (ESI) calc'd for $[C_{14}H_8Cl_2FN_4O_2]^-$([M-H]$^-$): *m/z* 353.0, found 353.0.

2.7 References

1. Semenza, G. L., *Nature Rev.* **2003**, *3*, 721–732. Targeting HIF–1 for cancer therapy.

2. Nagle, D. G., Zhou, Y–D. *Curr. Drug Targets.* **2006**, *7*, 355–369. Natural product–based inhibitors of hypoxia–inducible factor–1 (HIF–1).

3. Chang, H., Shyu, K–G., Lee, C–C., Tsai, S–C., Wang, B–W., Lee, Y. H., Lin, S., *Biochem Biophys. Res. Commun.* **2003**, *302*, 95–100. GL331 inhibits HIF–1a expression in a lung cancer model.

4. Cao, Z., Fang, J., Xia, C., Shi, X., Jiang, B. H., *Clin. Cancer Res.* **2004**, *10*, 5253–5263. *trans*–3,4,5'-Trihydroxystibene inhibits hypoxia–induction factor 1a and vascular endothelial growth factor expression in human ovarian cancer cells.

5. Mohammed, K. A., Hossain, C. F., Zhang, L., Bruick, R. K., Zhou, Y. D., Nagle, D. G., *J. Nat. Prod.* **2004**, *67*, 2002–2007. Laurenditerpenol, a new diterpene from the tropical marine alga *Laurencia intricate* that potently inhibits HIF–1– mediated hypoxic signal in breast tumor cells.

6. Rogers, J. L., Bayeh, L., Scheuermann, T. H., Longgood, J., Key, J., Naidoo, J., Melito, L., Shokri, C., Frantz, D. E., Bruick, R. K., Gardner, K. H., MacMillan, J. B., Tambar, U. K., *J. Med. Chem.* **2013**, *56*, 1739–1747. Development of inhibitors of the PAS–B domain of the HIF–2α transcription factor.

7. Lakshmaiah, G., Kawabata, T., Shang, M., Fuji, K., *J. Org. Chem*. **1999**, *64*, 1699-1704. Total synthesis of (–)-Horsfiline via asymmetric nitroolefination.

8. Leonard, K. A., Hall, J. P., Nelen, M. I., Davies, S. R., Gollnick, S. O., Camacho, S., Oseroff, A. R., Gibson, S. L., Hilf, R., Detty, M. R., *J. Med. Chem.* **2000**, *43*, 4488–4498. A selenopyrylium photosensitizer for photodynamic therapy related in structure to the antitumor agent AA1 with potent in vivo activity and no long–term skin photosensitization

9. Smallheer, J. M., Pinto, D. J., Wang, S., Qiao, J. X., Han, W., Hu, Z., Patent: US 20040006062 A1 20040101. Preparation of sulfonylaminovalerolactams as factor Xa inhibitors.

10. Teuber, L., Larsen, J. S., Patent Int. Appl. WO 2004087137 A1 20041014. A preparation of benzimidazole derivatives and their use for modulating GABAA receptor complex.

11. Abdurazakov, A. Sh, Elmuradov, B. Zh., Turdibaev, Zh. E., Shakhidayotov, Kh. M., *Chemistry of Natural Compounds*. **2009**, *45*, 402-408. Reaction of 2,3-tetramethylen-3,4-dihydroquinazol-4-one and its derivatives with aromatic aldehydes and furfuol,

12. Galan, A. A., Chen, J., Du, H., Forsyth, T., Huynh, T. P., Johnson, H. W. B., Kearney, P., Leahy, J. W., Lee, M. S., Mann, G., Patent WO2008/42282. Preparation of 1H–imidazole–4,5–dicarboxamides as JAK–2 modulators.

13. Keana, J. F. W., Kher, S. M., Cai, S. X., Dinsmore, C. M., Glenn, A. G., Guastella, J., Huang, J–C., Ilyin, V., Lu, Y., *J. Med. Chem.* **1995**, *38*, 4367–4379. Synthesis and structure–activity relationships of substituted 1,4–dihydroquinoxaline–2,3–diones: antagonists of N–methyl–D–aspartate (NMDA) receptor glycine sites and non–NMDA glutamate receptors.

14. Saha, B., Sharma, S., Sawant, D., Kundu, B., *Tetrahedron*, **2008**, *64*, 8676. Application of the Pictet–Spengler reaction to aryl amine substrates linked to deactivated aromatic heterosystems.

15. Ray, S., Patra, A., Mal, D., *Tetrahedron,* **2008**, *64*, 3253-3267. Tandem annulation strategy for the convergent synthesis of benzonapthopyranones: total synthesis of chartarin and O-methylhayumicinone.

16. Kristensen, J. L., *Tetrahedron Letters*, **2008**, *49*, 2351-2354. Synthesis of RHPS4 via an anionic ring-closing cascade.

CHAPTER THREE

IDENTIFICATION OF A NATURAL PRODUCT INHIBITOR OF HIF

3.1 A marine–bacterial derived natural product library for high- throughput screening.

In parallel with the high–throughput AlphaScreen to identify synthetic inhibitors of HIF-2α, the 1200 natural product fractions of the MacMillan lab library were also screened. The results from this screen resulted in the discovery of an oxysterol natural product, which has opened up a new avenue of research towards the endogenous ligand of the HIF transcription factor. Those results are described here.

The primary goal of research in the MacMillan lab is the discovery and characterization of natural products from marine–derived sources such as hunanamycin A, an antibacterial natural product with a pyrido[1,2,3-de]quinoxaline-2,3-dione motif[1] as well as a series of ammosamide analogs with moderate cytotoxicity activity[2] (**Figure 3.1**).

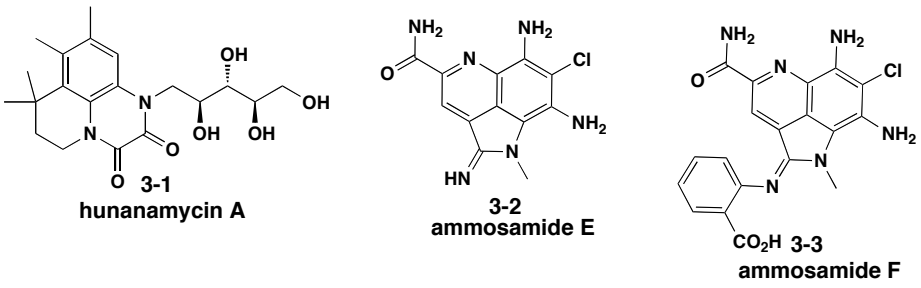
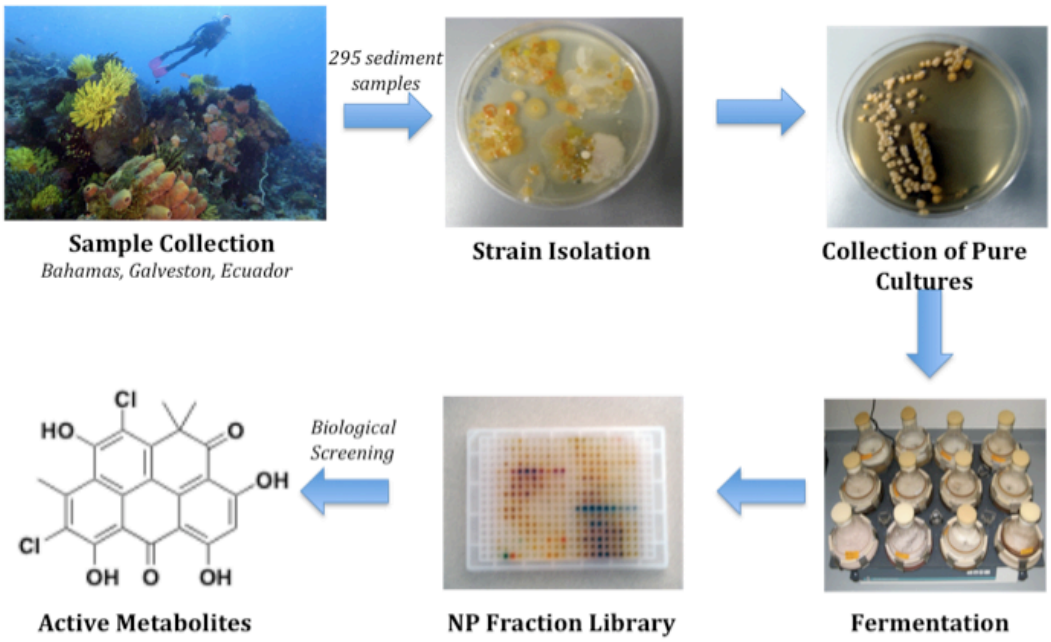


Figure 3.1 Natural products isolated in the MacMillan lab

To do this, our lab collects sediment samples and marine sponges in tropical marine environments such as the Bahamas (**Scheme 3.1**). These samples are brought back to the lab, where we work to cultivate the bacteria present in the samples. The sediment samples are dried and crushed using a mortar and pestle. A small circular sponge is dipped in the dry sediment sample and then plated onto an agar plate, using a variety of media conditions. Bacteria of distinct morphologies are selected and re–plated until pure bacterial cultures are isolated.



Scheme 3.1 Process of natural product isolation in the MacMillan lab

To generate the natural product fraction library, each culture is grown in liquid media (5 x 1.0 L). After seven days, XAD–7, a lipophilic resin is added to the cultures to absorb the natural products. The resin is collected and subsequently soaked overnight in acetone to give a crude extract. Crude samples are fractionated via automated C18 chromatography (ISCO) to

give 20 fractions, each containing on average 1-10 compounds. These fractions make up a natural product library, which is available to the in-house high–throughput–screening facility for screening in biological assays

At the time of screening in the HIF–2α AlphaScreen assay, the library contained 1200 fractions. Utilizing the AlphaScreen we identified a single fraction that had an appreciable ability to inhibit the HIF–2α PAS–B/ARNT heterodimerization. The fraction was identified as SNA014-6.

3.2 Identification and description of SNA-014

As mentioned above, each bacterial crude extract is purified via automated separation on a C-18 column with gradient elution to give 20 fractions. This purification crudely separates out the organic compounds and helps to remove undesirable salts from the mixture. Each fraction contains on average 1-10 organic compounds. AlphaScreen data showed biological activity primarily in fraction 6, SNA-014-6. Bacterial strain SNA-014 was isolated from a sediment sample collected in a lagoon at Little San Salvador, Bahamas in 2008 and cultivated using A1 media composed of yeast, peptone, and starch and also containing cyclohexamide. SNA-014 was phylogenetically characterized using 16S rRNA to show 98% identity to the nontraditional actinomycete *Gordonia polyisoprenivorans*. The material from fraction six was further purified via RP–HPLC on a Phenomenex Luna C18 column with gradient elution. Again the activity of the fractions was evaluated in the AlphaScreen in which fraction six showed activity (SNA-014-6-6). In addition, a dose response curve further validated this hit (**Figure 3.2**).

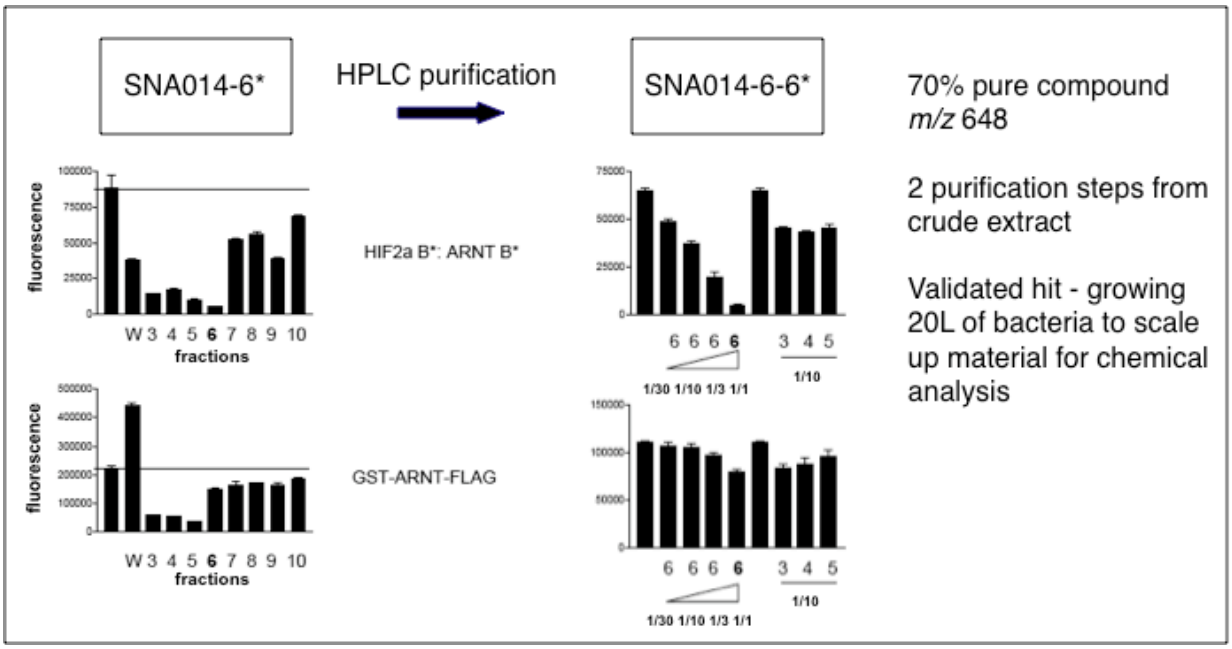
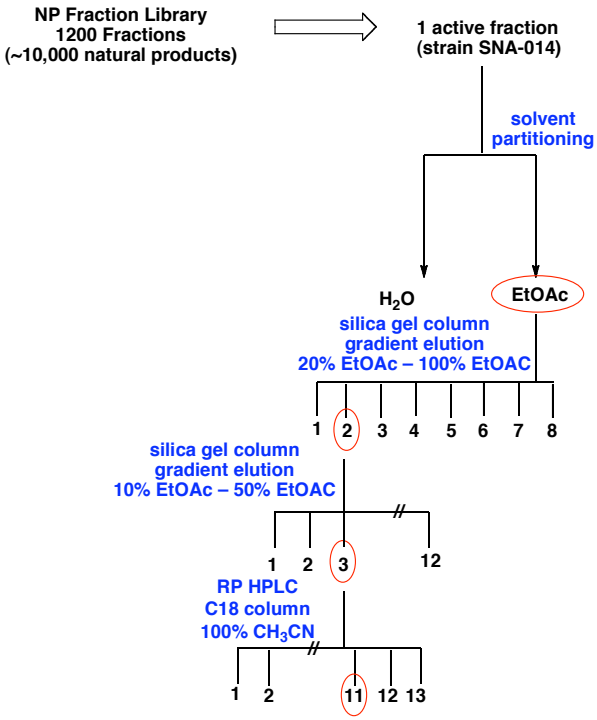


Figure 3.2 Biological activity of initial SNA-014 fraction

In order to characterize the active component of this crude mixture, a large–scale fermentation (20.0 L) of SNA-014 was grown. This bacteria was difficult to culture in our standard A1 media so TCG media containing tryptone, casitone, and glucose was selected. This provides a different source of carbon and nitrogen for the organism. The strain was grown in TCG media at 32 ºC for 14 days. Again, the excreted metabolites were collected using XAD-7 resin and the resulting crude extract was purified by an assay-guided fractionation. The crude organic material was partitioned between water and ethyl acetate. The ethyl acetate fraction was further purified via flash chromatography on silica gel with a gradient elution from 20% ethyl acetate in hexane to 100% ethyl acetate. Eight fractions were collected and analyzed via AlphaScreen, which showed activity in fraction two. This fraction was further purified via flash chromatography on silica gel

with a gradient elution from 10% ethyl acetate in hexane to 50% ethyl acetate in hexane. Twelve fractions were collected with the desired activity identified in fraction three. Fraction three was purified via RP-HPLC on a C18 column with a standard elution of 100% CH_3CN for 35 min. Thirteen fractions were collected and evaluated via AlphaScreen. The desired activity was identified in fraction eleven (**Figure 3.3**). This fraction was evaluated via ^{1}H and ^{13}C NMR.

Figure 3.3 Purification of SNA-014-E-2-3-11



3.3 Isolation and characterization of a natural product inhibitor of HIF-2.

Initial 1D NMR experiments (^{1}H, ^{13}C NMR) of the natural product revealed the structure was an oxy–sterol molecule (**Figure 3.4**). This is observed through the large number of aliphatic signals in the 0.5 to 2.5 ppm range as well as the olefin proton shown at 5.3 ppm and the

oxygenated methine proton at 3.5 ppm. In addition, proton NMR showed at least six methyl groups, two of which were from an isopropyl group.

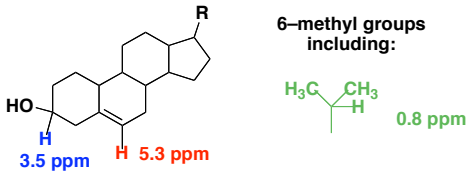


Figure 3.4 Partial structure determination of SNA-014-E-2-3-11

To further evaluate the structure of the sterol compound, a series of 2D NMR experiments were used. This included: COSY to show three bond $^{1}H–^{1}H$ correlations, HMBC to show three and four bond $^{1}H–^{13}C$ correlations, HSQC to show one bond $^{1}H–^{13}C$ correlations. Unfortunately, the structure of the molecule was difficult to elucidate even with the 2D NMR data. The majority of the NMR peaks showed up between 0.5 and 1.5 ppm in the ^{1}H NMR. Similar overlap was observed in the ^{13}C NMR spectrum. To help elucidate individual spin systems of the molecule a selective 1D TOCSY experiment was utilized. TOCSY allowed us to irradiate a single proton within the molecule and the spectrum produced showed only ^{1}H NMR peaks from within the same spin system. This let us draw out individual peaks from the heavily overlapping region of the spectrum.

With the TOCSY data we were able to identify an aliphatic spin system based on the alcohol–substituted methine proton at 3.50 ppm (**Figure 3.5**). We continued to irradiate individual protons, but unfortunately, every proton that we could isolate was a part of the same

spin system that had already been identified. Because of the overlap we were not able to separate out additional spin systems.

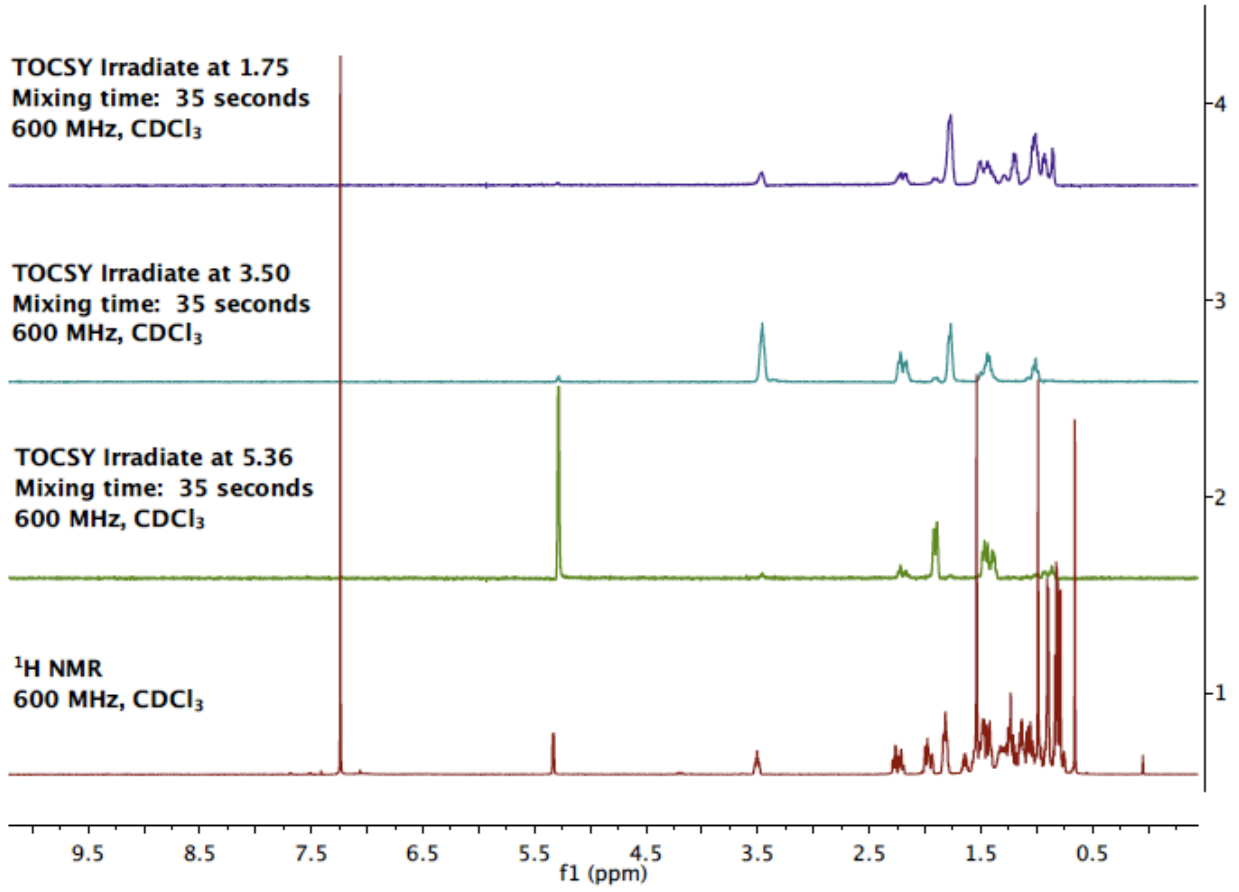


Figure 3.5 TOCSY experiment to identify individual spin systems

Based on the existing data, we knew that the natural product was a hydroxy sterol. However, using traditional NMR methods, it was not possible to determine the complete structure of the molecule. We also realized that a hydroxy sterol molecule was not a good candidate for a medicinal chemistry campaign, but could provide valuable clues for a potential endogenous ligand.

3.4 Investigation into the endogenous ligand of HIF–2α

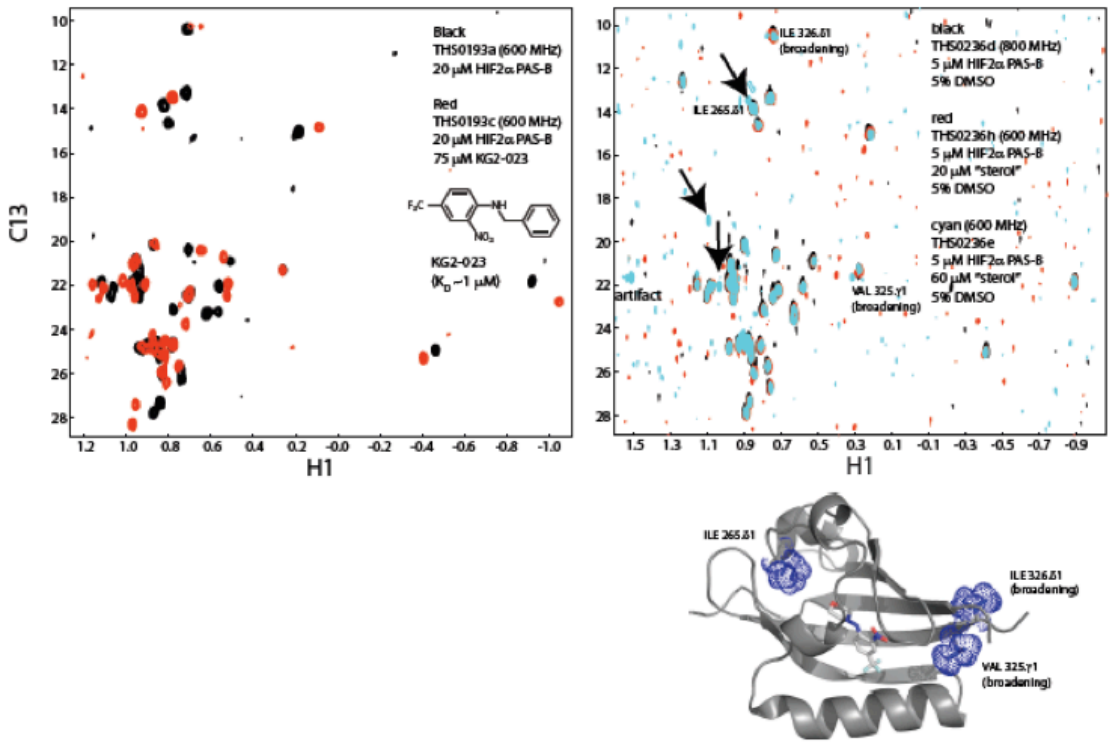


Figure 3.6 NMR data confirming sterol interaction with internal pocket of HIF–2α PAS–B

Although this molecule was not a good medicinal chemistry target, we knew that the PAS–B domain of HIF–2α contained a large pocket and this suggested that an endogenous ligand might be responsible for direct regulation of HIF. This natural product was further evaluated as a ligand of HIF–2α transcription factor. To look at direct interaction of the natural product with the protein, the NMR assay using ^{13}C labeled protein was performed. The lipid molecule was poorly soluble in the buffer in which this binding assay was performed, and therefore DMSO (5 %) was added to the solution. The poor solubility limited the concentration

of the natural product that could be obtained in this experiment. However, characteristic shifts in the NMR confirmed that the molecule was in fact binding within the hydrophobic pocket of the protein in a similar fashion to what was observed with the small molecules in the previous chapter (**Figure 3.6**).

3.5 Discussion

The data from these assays confirmed that our natural product behaved similarly to the other HIF inhibitors. Although this molecule does not possess ideal characteristics as a therapeutic or chemical probe in further HIF studies, this discovery was still interesting. What intrigued us about this new HIF ligand was that sterols are prevalent in many biological systems and perhaps this could point to an endogenous ligand of the transcription factor.

As more becomes understood about PAS domains, studies have shown that these domains are often regulated by natural and artificial ligands. These ligands act through allosteric mechanisms to alter the conformation of the protein and interrupt protein–protein interactions. With the identification of a sterol ligand for these transcription factors, we began to investigate if the natural product may also be related to the endogenous ligand responsible for regulation of the protein.

Interestingly, studies are starting to show a strong connection between sterol biosynthesis and the hypoxia response pathway. In yeast, sterol biosynthesis is directly regulated by oxygen availability. When oxygen levels are low, the transcription factor sre1 is activated. This in turn activates sterol biosynthetic enzymes. Sre1 is also responsible for turning on genes that allow the organism to adapt to hypoxic conditions.[3] In mammals the opposite connection between sterol biosynthesis and hypoxia is observed. Hypoxia results in the down–regulation of sterol

biosynthesis (**Figure 3.7**).[4] The connection between hypoxia and sterol biosynthesis made our discovery of a sterol as an inhibitor of HIF heterodimerization that much more interesting. We wanted to see if a sterol might be the endogenous ligand of the HIF–2α PAS–B domain.

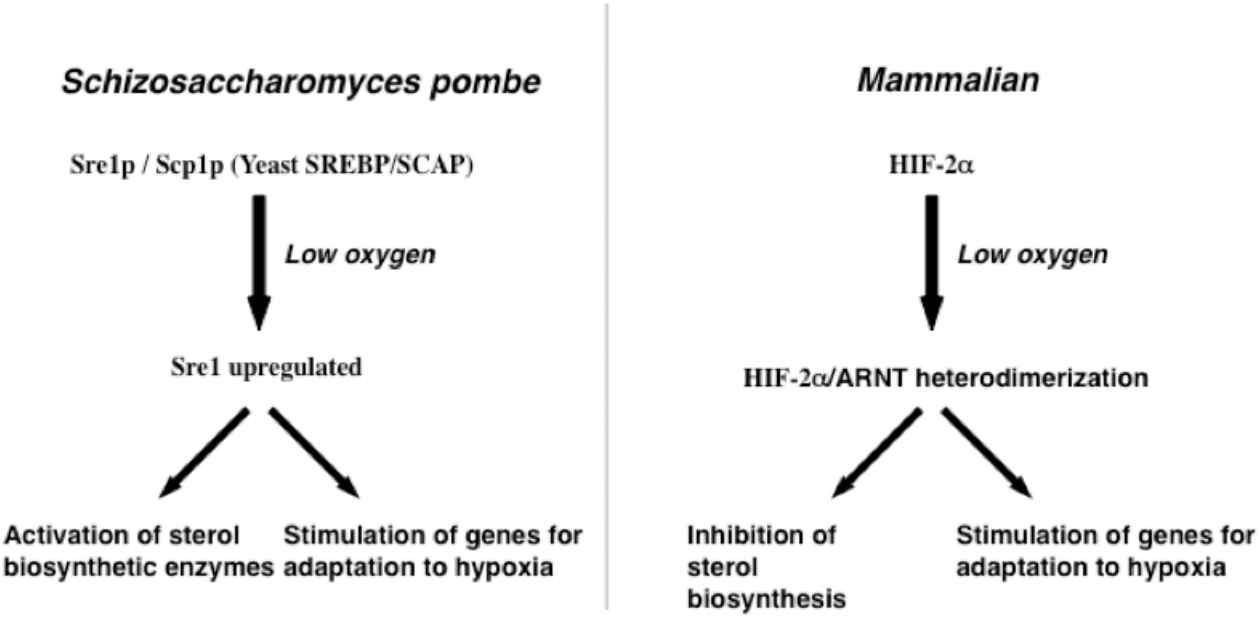


Figure 3.7 The role of sterols in the hypoxic response pathway

There are two primary approaches that we have pursued to identify the endogenous ligand of the HIF–2 transcription factors. The first approach is the candidate approach. A series of commercially available sterol derivatives were screened in the AlphaScreen and the NMR assay to evaluate the efficacy of these molecules (**Figure 3.8**).

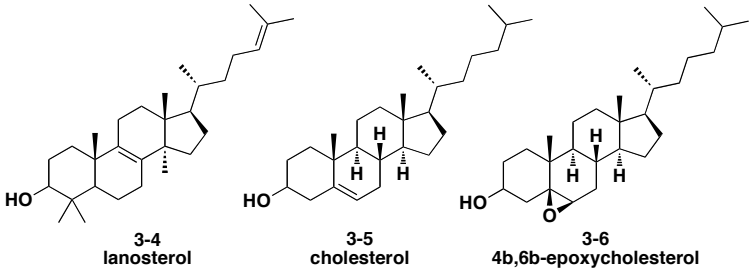


Figure 3.8 Commercial sterols tested for HIF inhibition

The second approach involved purifying active natural products from mammalian tissue to identify a ligand originating from mammalian sources. Both of these approaches are currently being pursued by members of our lab and other labs within the biochemistry department at UT Southwestern.

3.6 Conclusion

Through the AlphaScreen of approximately 10,000 secondary metabolites led to the identification of a single marine–bacterial derived natural product. While the structure determination of this molecule was challenging, the identification of a sterol ligand for the HIF–2 transcription factor has led to a better understanding of the identity of the endogenous ligand of this protein. Studies are ongoing to further identify the endogenous ligand in this pathway.

3.7 References

1. Hu, Y., Wang, K., MacMillan, J. B., *Org. Lett*. **2013**, *15*, 390–393. Hunanamycin A, an antibiotic from a marine–derived bactillus hunaanensis.

2. Pan, E., Oswald, N. W., Legako, A. G., Life, J. M., Posner, B. A., MacMillan, J. B., *Chem. Science*. **2013**, *4*, 482–488. Precursor–directed generation of amidine containing ammosamide analogs: ammosamides E–P.

3. Hughes, A. L., Todd, B. L., Espenshade, P. J. *Cell*. **2005**, *120*, 831. SREBP Pathway responds to sterols and functions as an oxygen sensor in fission yeast.

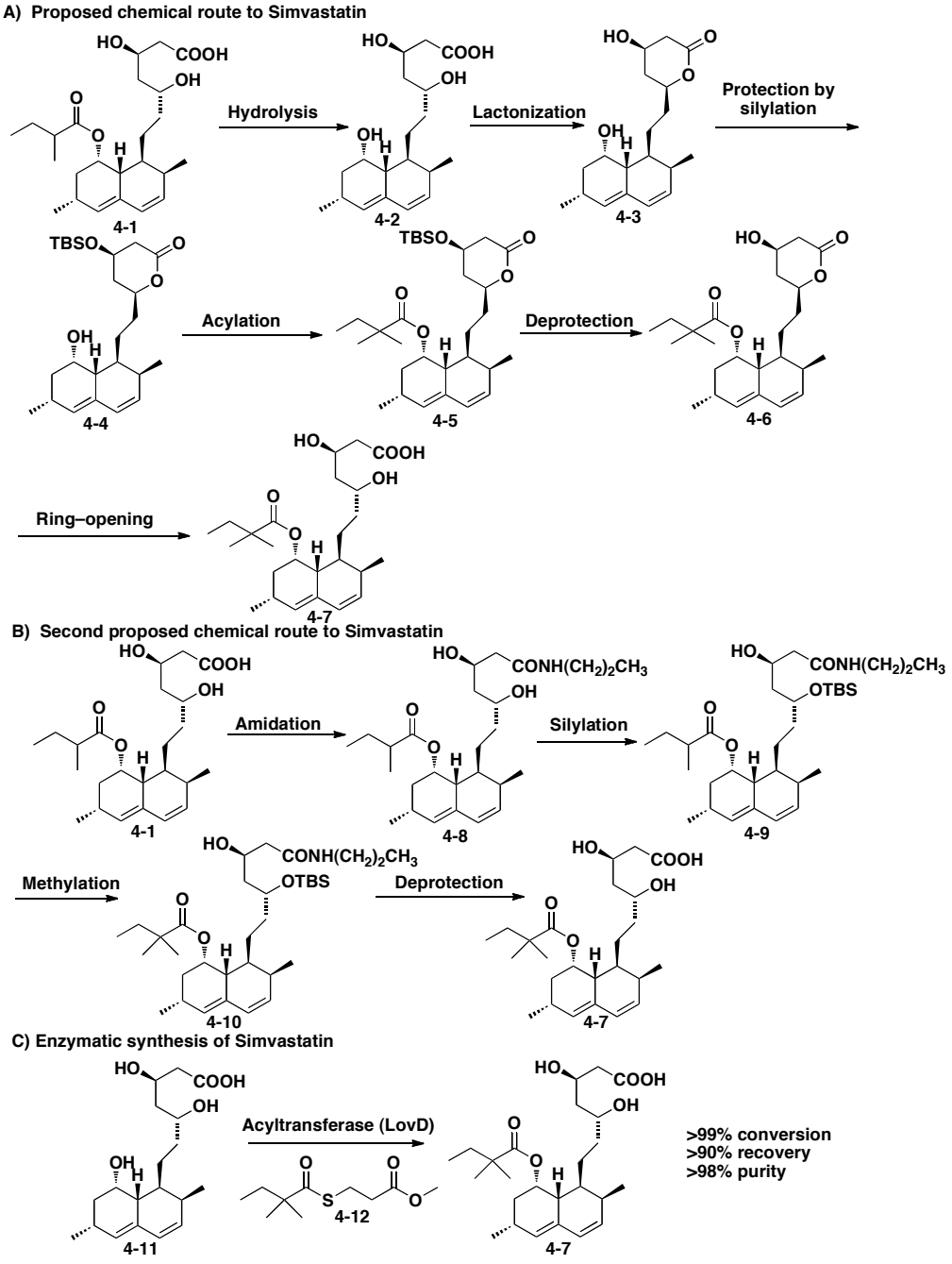
4. Nguyen, A. D., McDonald, J. G., Bruick, R. K., Debose–Boyd, R. A., *J. Chem. Biol*. **2007**, *282*. 27436. Hypoxia stimulates degredation of 3–hydroxy–3–methylglutaryl–coenzyme A reductase through accumulation of lanosterol and hypoxia–inducible factor–mediated induction of insigs.

CHAPTER FOUR

NEW REACTION DISCOVERY: FROM BIOCATALYSIS TO TRANSITION METAL CATALYSIS

4.1 History of biocatalyzed reactions in organic synthesis

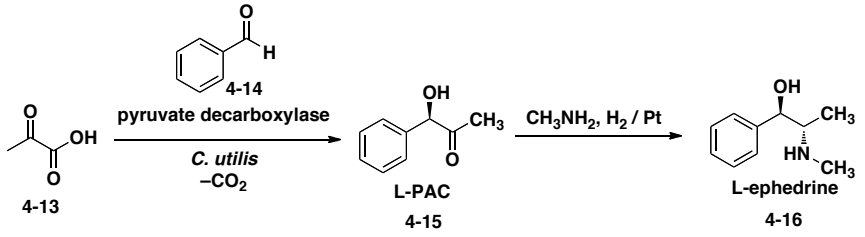
The development of new catalysts for organic transformations is at the forefront of research in organic chemistry. Over the past decade, significant progress has been made in the area of transition metal catalysts. Cross coupling reactions and olefin metathesis are two important examples where transition metals have been used for the development of new chemistry. However, biocatalysis is also an important tool for organic chemists for several reasons. Reactions catalyzed by biological enzymes have the advantage of regio- and stereo-control while utilizing mild reaction conditions. Specifically, many biocatalyzed transformations can be performed in water at ambient reaction temperatures. In fact, the synthesis of Simvastatin, a cholesterol–lowering drug, illustrates an excellent example of some of the benefits that biocatalysis can provide (**Scheme 4.1**). Three approaches towards the synthesis of Simvastatin have been developed.[1] Two of these routes are via traditional synthetic methods, which require several protection/deprotection steps. This reduces the overall synthetic efficiency. However, the third approach utilizes an acyltransferase enzyme, which allows for a one–step conversion. The desired product is obtained in high yield and purity. This example highlights one of the major benefits of biotransformations, selectivity.



Scheme 4.1 Chemocatalytic and enzymatic synthesis of Simvastatin[1]

4.1.1 Early report of biocatalysis in industrial applications

Biocatalyzed organic transformations have a rich history within the pharmaceutical industry. One of the first reported uses of a biocatalyzed reaction towards the synthesis of a pharmaceutical compound was for the synthesis of pseudophedrine, used as a decongestant. Pseudophedrine is a naturally occurring alkaloid produced by the plant species Ephedra. However, a more efficient route to obtaining this compound involves the biocatalyzed conversion of dextrose and benzaldehyde to a pseudophedrine precursor, L–PAC. This is then followed by chemical conversion to the desired compound using reductive amination (**Scheme 4.2**). As pseudophedrine is one of the most common ingredients in many cold medicines, this process is important on a multi-ton scale and biocatalysis contributes to its efficient production.

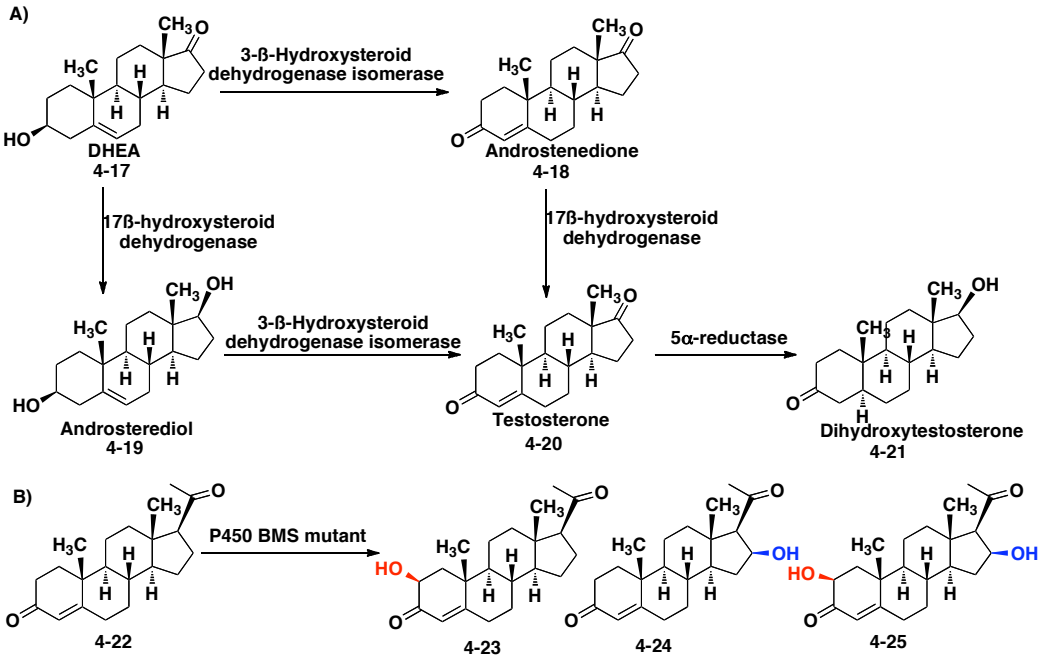


Scheme 4.2 Biocatalyzed synthesis of pseudophedrine[2]

4.1.2 Vast applications of biocatalysis in manufacture of sterols, past and present.

Another important example of biocatalysis in the pharmaceutical industry is the production of functionalized sterols for multiple purposes including; wound healing medicines, performance enhancers, and hormone treatments.[3] Sterols represent a group of relatively unfunctionalized molecules. Synthetic alterations to the un–activated carbons of the sterol framework are difficult using standard synthetic reagents, due to the difficulty in discriminating

nearly equivalent sp^3 carbons. The selectivity that enzymes can display has made them invaluable for synthetic modifications of the sterol framework. As such, naturally occurring enzymes are used in the partial synthesis of a variety of useful sterols (**Scheme 4.3a**). Through "laboratory evolution" enzyme selectivity can be tailored to specific oxidation selectivity. This is illustrated in **Scheme 4.3b** where the oxidizing enzyme cytochrome P450 was mutated at specific residues. This changed the specificity of the enzyme and allowed the selective oxidation at either the 2-position (**4-23**), the 16–position (**4-24**), or both (**4-25**). Biocatalysis remains an important tool in the synthesis of many steroid molecules.

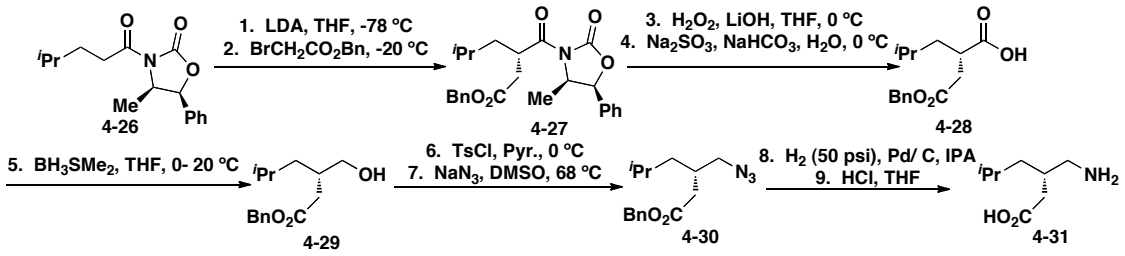


Enzyme with mutation	% 4-23	% 4-24	% 4-25
Cytochrome P450 with F87A Mutation	18	82	0
Cytochrome P450 with F87A and A330W Mutation	9	91	0
Cytochrome P450 with F87A and V78V and A82N Mutation	100	0	0
Cytochrome P450 with F87A, R47Y, T49F, V78L, and A82M Mutation	11	64	22

Scheme 4.3 Biocatalysis with steroid production.[3-5]

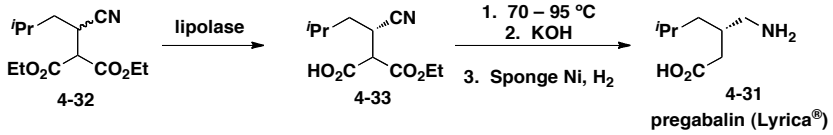
4.1.3 Development of efficient Lyrica® synthesis utilizing biocatalysis

Along with oxidation reactions, biocatalyzed kinetic resolutions reactions are also very important in pharmaceutical synthesis as we see in the total synthesis of Lyrica® by Pfizer Global Pharmaceuticals. The initial synthetic route to Lyrica® was relatively inefficient for this fairly simple molecule (**Scheme 4.4**).[6]



Scheme 4.4 Original synthetic route to Lyrica

This process involved the use of an expensive chiral auxillary to set the absolute stereochemistry in the molecule. In addition, extreme temperatures and the use of a hazardous azide reagent (steps 1 and 7, respectively) make this process unattractive for large–scale production. Pfizer was able to develop an alternative biotransformation strategy to Lyrica that involved a kinetic resolution of racemic diester starting material (**4-32**) using a lipolase enzyme. The desired compound was then obtained efficiently in three additional steps (**Scheme 4.5**).



Scheme 4.5 Biocatalytic synthesis of Lyrica[7]

In this example, the biocatalyzed route offered several advantages over the initial synthesis including fewer steps, increased yield, lower energy usage and lower overall cost of production.[7] In fact, several pharmaceutical compounds currently available on the market are synthesized in part through biocatalysis. A partial list of these compounds is illustrated in **Figure 4.1**.

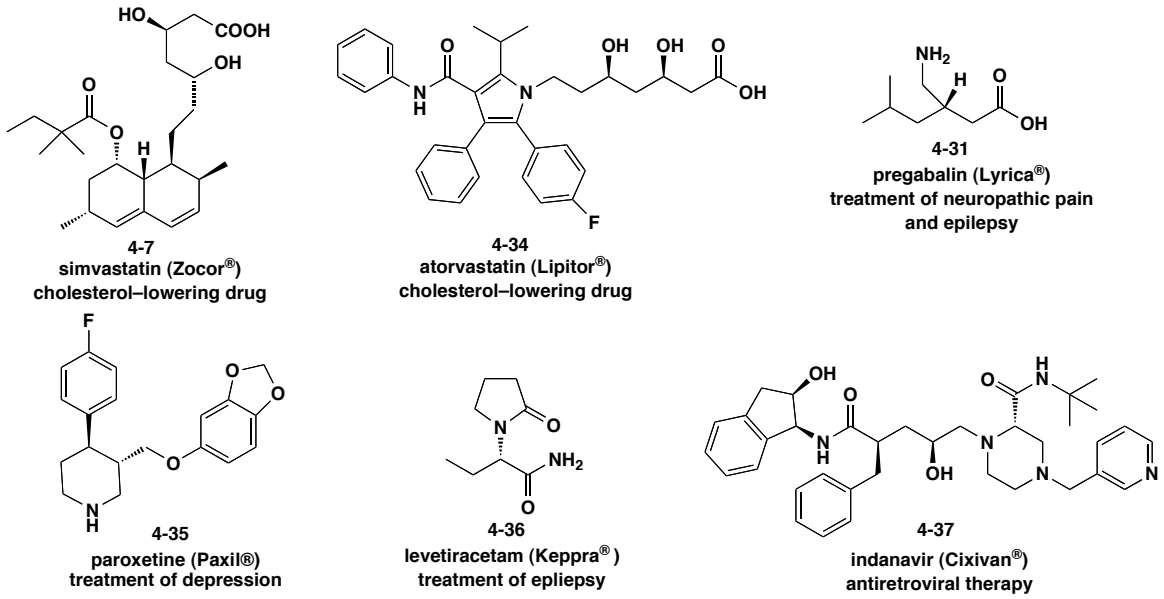


Figure 4.1 Partial list of pharmaceutical compounds produced in part through biosynthesis[7]

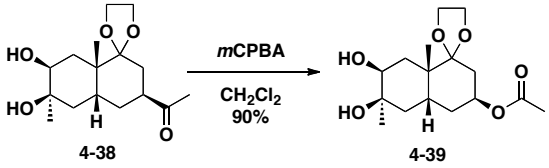
4.2 Baeyer–Villiger oxidation: one of the best-studied biocatalyzed reactions

Although biocatalyzed reactions can be very important in the synthesis of various molecules, there can also be several drawbacks to their development. Enzymes can have limited stability in the reaction conditions and this can make them difficult to work with.[8] Another limitation is that enzymes have inherent specificity. While this can lead to remarkable stereo– and regio–selectivities, this can also lead to poor substrate scope. Many of the complications in

using enzymes for organic transformations can be overcome through a variety of techniques that will be discussed later. However, these limitations have made efforts to discover and understand *new* biocatalyzed reactions fairly minimal. Current efforts in the field are typically focused on expanding the utility of known enzymes such as lipases, esterases, etc., which are already well understood. Fewer efforts are expanded towards the discovery of new enzymatic reactions. As an example, one of the most studied families of enzymes for biocatalysis is the Baeyer-Villiger monooxygenase (BVMO), and this enzyme continues to be highly studied today.

4.2.1 The traditional non–enzymatic Baeyer– Villiger oxidation

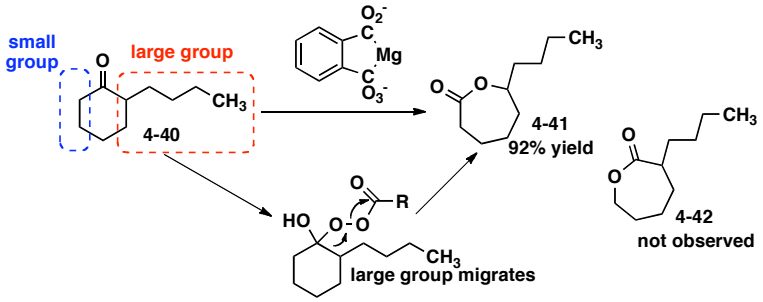
The Baeyer- Villiger oxidation describes a reaction in which a ketone is converted to the corresponding ester or lactone. Using traditional methods this reaction is performed using a peroxy acid, such as *m*CPBA, often in the presence of a Lewis acid catalyst. For example, in the synthetic route towards the synthesis of taxol, the highly functionalized methyl ketone containing decalin system was treated with *m*CPBA to introduce oxygen adjacent to the ring (**Scheme 4.6**).



Scheme 4.6 Non–enzymatic Baeyer–Villiger oxidation towards the synthesis of taxol.[9]

The non–enzymatic version of the Baeyer–Villiger reaction can be high yielding and it is tolerant of a variety of functional groups, as described in the previous example. However, there are important advantages to the enzymatic version of this reaction. For example, in the non–enzymatic reaction, the regioselectivity is dictated by the sterics of the migrating group. The

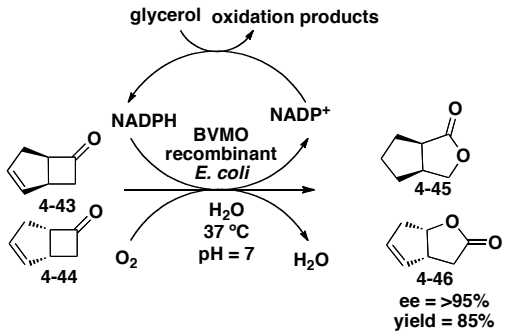
larger group of the ketone will migrate following the oxidation, leading to the regiospecific ester product (**Scheme 4.7**). The regioselectivity, therefore, is dependant on the degree of difference in steric bulk between the two ketone substituents.



Scheme 4.7 Regioselectivity of the Baeyer–Villiger oxidation[10]

4.2.2 Identification of Baeyer–Villiger monooxygenase enzymes

In biological systems the Baeyer–Villiger oxidation reaction can be catalyzed by a group of enzymes known as Baeyer–Villiger Monooxygenases (BVMO). BVMOs are B flavoprotein monooxygenases that utilize NADPH as an electron donor. As a proof–of–principle experiment on the scalability of the BVMO transformation, it was applied on an industrial scale towards the oxidation of racemic bicyclo[3.2.0]hept-2-en-6-one (**Scheme 4.8**).



Scheme 4.8 Industrial scale use of BVMO[13]

This class of enzymes has been shown to perform the Baeyer–Villiger oxidation reaction on a variety of substrates with high regio– and stereo–selectivity. In natural settings, the BVMO enzyme is important to catabolism by allowing microorganisms to utilize various ketones as a food source.[11] These enzymes are also important in the biosynthesis of several secondary metabolites (**Figure 4.2**).

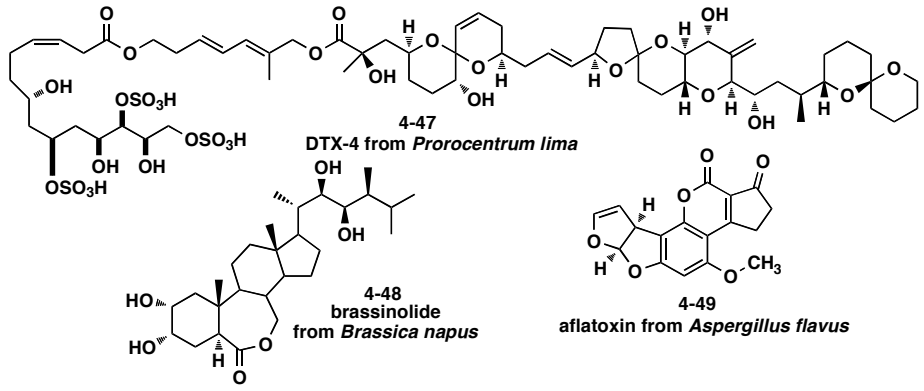
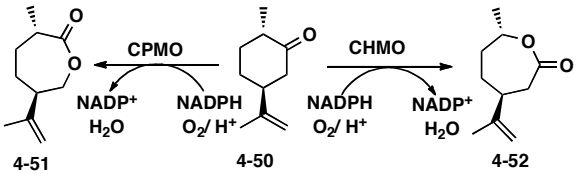


Figure 4.2 Natural products whose biosynthesis involves the Baeyer–Villiger oxidation.[12]

As stated earlier, an advantage to the enzymatic B–V oxidation is that high regioselectivity can be obtained. As we saw in the non–enzymatic reaction, the regioselectivity is controlled by the sterics of the molecule. A useful characteristic of enzymatic oxidations is that it is often possible to alter the regioselectivity by changing the enzyme (**Scheme 4.9**) or by changing the isomer of the starting material. In the example illustrated below, two different versions of the BVMO are used. CPMO (cyclopentanone monooxygenase) and CHMO (cyclohexanone monooxygenase) reference specific substrate selectivity of BVMO enzymes. This specificity can also be parlayed into regioselectivity as this example illustrates.

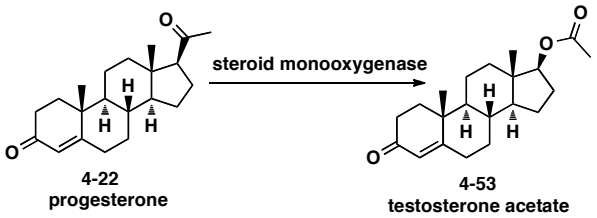


Scheme 4.9: Regioselective biocatalyzed Baeyer–Villiger oxidation.[13]

The enzymatic reaction also works with moderate reaction temperatures and the use of water as the solvent. As a result, this oxidation is very attractive for industrial processes. The important utility of this reaction has lead to a breadth of research into the optimization of the enzymatic transformation.

4.2.3 Optimization efforts towards enzymatic Baeyer–Villiger oxidation

As the Baeyer–Villiger monooxygenase class of enzymes represents some of the most studied enzymes, they are good models for understanding how biocatalysis research is focused. Although the non- enzymatic Baeyer–Villiger oxidation was first reported in the 19th century, the first hint at an enzymatic version of the Baeyer–Villiger oxidation likely occurred in 1948 when Turfitt described steroid degradation in fungal cultures. Later, in the 1960's an enzymatically catalyzed lactonization of a sterol was observed.[14] It wasn't until 1986 that Trudgill and Itagaki were able to purify a steroid monooxygenase using affinity chromatography. They demonstrated that the purified enzyme was able to catalyze the oxidation of progesterone to testosterone acetate (**Scheme 4.10**).



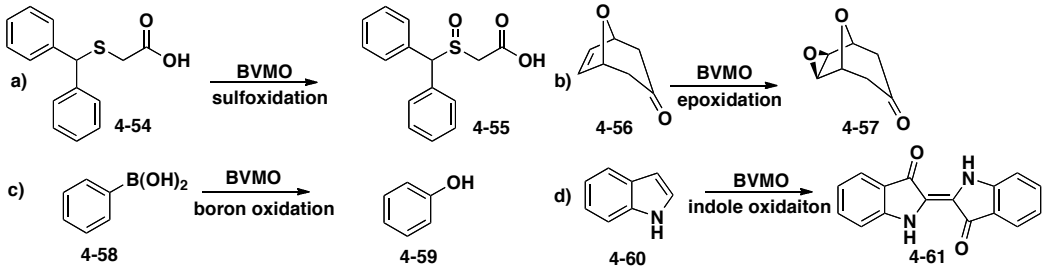
Scheme 4.10: Purified steroid monooxygenase able to perform the Baeyer–Villiger reaction.

While Baeyer–Villiger monooxygenase activity was first observed in the 1940's, the genes that encode this enzyme were not described until the 1990's.[11] This breakthrough was important for several reasons. It allowed scientists to mine sequential microbial genomes in search of novel BVMO's. More importantly, it allowed scientists to clone the encoding gene into a heterologous host such as *E. coli.* By analyzing genome sequences of microorganisms, a plethora of BVMO's have been identified with at least one BVMO gene described in each microbial genome, and with the highest abundance of BVMO's found in actinomycetes and some filamentous fungi. Today research efforts in this area are focused on the indentification and purification of novel BVMO's, exploration of applications of existing BVMO's, as well as optimization of current BVMO's.

4.2.4 Various roles for BVMO's in organic synthesis

Baeyer- Villiger monooxygenase enzymes have been shown to catalyze not just the Baeyer- Villiger oxidation reaction but several other oxidation reactions as well (**Scheme 4.11**). In addition to the Baeyer- Villiger reaction, ester hydrolysis and saponification reactions are

catalyzed by lipase enzymes. Libraries of these enzymes can now be purchased so that one can screen for reactivity that matches their individual needs.



Scheme 4.11: Alternative uses of BVMO's.[11]

4.3 Methods of screening for new biocatalyzed reactions

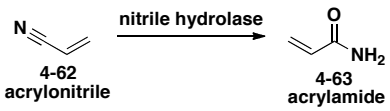
Currently, biocatalysis research is focused on the optimization of known biocatalyzed reactions: hydrolysis, esterification, and oxidation. Chemists have access to a toolbox of enzymes that can effectively catalyze useful oxidation and reduction reactions with specificity that is not available by traditional synthetic means. These are often sold as temperature stable powders that can be used as any other chemical reagent. However, there are often limitations in these enzymes, which may include instability of the enzyme, inhibition, low yield and low specificity. Through various engineering strategies, scientists have been able to optimize these reactions to make them more powerful tools to organic chemistry.

4.3.1 Screening for enzyme activity

While a relatively small number of classes of enzymes are available for synthetic applications, within these classes there is a large variety. For example, the lipase class of

enzymes is available from multiple sources including animals, plants and microorganisms. When a synthetic step requires a lipase enzyme, there are hundreds to choose from. In fact, libraries of enzymes are available from chemical vendors such as SigmaAldrich and chemists can screen an entire library to find the lipase that best matches their needs.

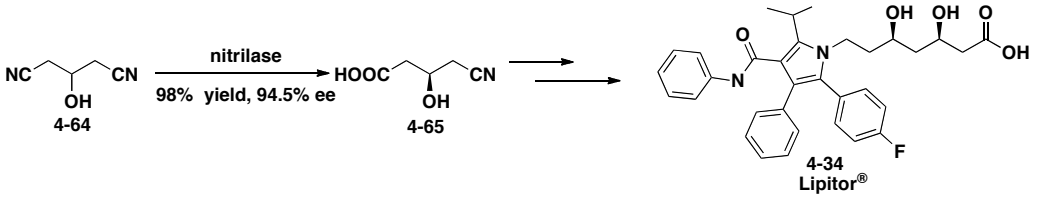
This screening process can also be manipulated to select for the best enzyme. This example is illustrated with the synthesis of acrylamide. Acrylamide is a simple molecule with important uses in many important industrial processes including the manufacture of many polymers.[15] This small molecule is produced on the industrial scale via hydrolysis of acrylonitrile by a nitrile hydrolase enzyme (**Scheme 4.12**).



Scheme 4.12: Biocatalyzed synthesis of acrylamide

4.3.2 Metagenome screening

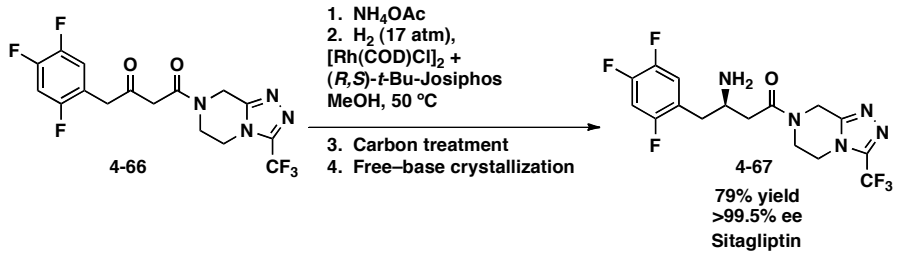
Another form of screening for effective enzymes is through metagenome screening. In this process DNA can be extracted from environmental samples and through the use of bioinformatics, specific genes can be identified for enzymes of interest. These genes can then be heterologously expressed in microorganisms, allowing enhanced production of the desired enzyme. This genome mining approach was useful in the discovery of 100 novel nitrilase enzymes.[15] A de-symeterization reaction of a dinitrile substrate was a key first step in the synthesis of Lipitor® by Pfizer Global Pharmaceuticals (**Scheme 4.13**).



Scheme 4.13: Nitrilase desymmeterization reaction *en route* to Lipitor®

4.3.3 Protein engineering

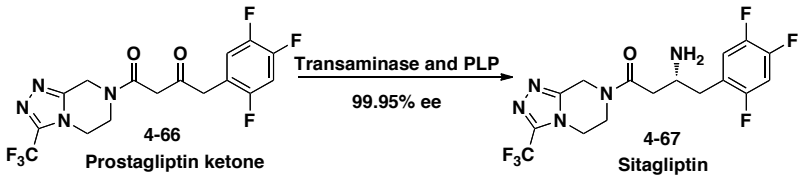
Sometimes when a reaction calls for a specific enzyme, the enzyme may not meet the requirements for that specific reaction. This can occur if the enzyme does not meet the substrate specificity needs, the regiospecificity needs, or does not provide the correct enantiomer. Through protein engineering it can be possible to alter the enzyme in a way to meet the needs of the desired reaction. An example of this process is illustrated in the synthesis of Sitagliptin, an antidiabetic drug.[8]



Scheme 4.14: Chemocatalytic synthesis of Sitagliptin

It was proposed that a transaminase enzyme could be used to afford Sitagliptin from the diketone precursor. However, initial screening approaches could not find an enzyme that could accommodate the bulky substrate. Through protein engineering, a transaminase from

Arthrobacter sp. was developed that could give the desired substrate specificity and afford the desired compound with high enantio–purity (**Scheme 4.15**).[8]



Scheme 4.15: Engineered transaminase to catalyze synthesis of Sitagliptin

4.4 Current state of whole–cell biocatalysis

While both isolated enzymes and whole cells are used for biocatalysis, there are some important benefits to the use of whole cell biocatalysis. Whole–cell biocatalysts require no purification of an enzyme, which reduces cost and expertise needed for use. Another benefit to whole–cell biocatalysis is that in the whole cell environment, the enzymes tend to be more stable as they are protected from the external environment. This can extend the life of the enzyme during the reaction. Finally, many enzymes require the use of cofactors, which are more easily recycled in the whole–cell environment.[16]

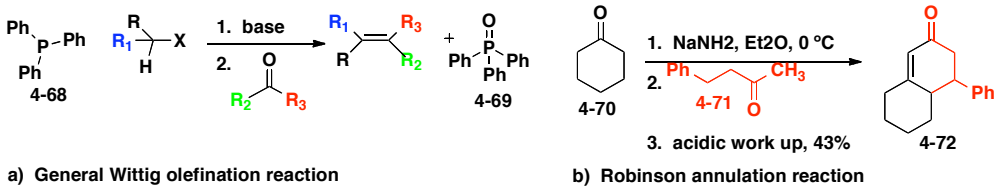
While whole–cell biocatalysis has important advantages, there are also some challenges specifically associated with the use of whole cells. For example, many microorganisms contain a multitude of active enzymes and this can interfere with the desired transformation. Specifically, if a microorganism contains many hydrolase enzymes with varying specificities, competition between enzymes can lead to poor selectivity. Recent advances in the field have been able to combat some of the disadvantages afforded by whole–cell use to improve on the wide–spread

application in synthetic chemistry. For example, transfection allows a specific enzyme to be expressed in a robust organism such as *E. coli*. The transfer of a single gene eliminates the presence of competing enzymes in the new cell. Also, protein engineering approaches allow over–expression of specific enzymes to greatly improve on their activity within a cell.

These approaches represent the current state–of–the–art for whole–cell biocatalysis. This also illustrates that the field is limited to working with known enzymes and not enough focus is put on exploring for new enzymatic transformations.

4.5 The discovery of new reactions in organic synthesis

It is clear that biochemistry continues to play an important role in synthetic chemistry, but this technology has not been fully exploited. It is necessary to find new biocatalysis reactions that can be utilized in synthetic chemistry. For synthetic chemists, the challenge of new reaction discovery has been a constant challenge. As such, chemists have been developing new ways of performing organic transformations for more than a century. Sometimes the discovery of new reactions occurs accidentally such as the case of the Wittig reaction (**Scheme 4.16a**). While this reaction was discovered serendipitously, the Wittig olefination remains one of the most important routes to new carbon–carbon double bond formation. Often, the discovery of a new reaction occurs out of necessity. For example, as Robinson and co-workers approached the synthesis of sterols they needed to be able to prepare substituted cylcohexenones. They developed what is now known as the Robinson annulation. This is an example of new reaction discovery that came about out of necessity.

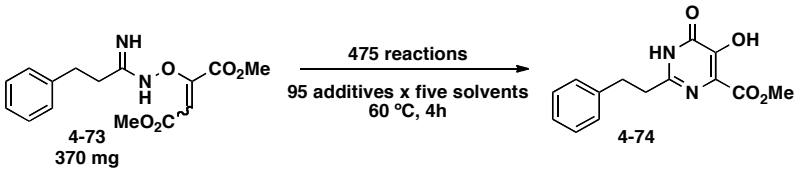


Scheme 4.16 Important name reactions[17]

Scientific imagination allows us to think of countless possibilities within organic synthesis. However, we can often be limited by our imaginations or more importantly, our knowledge of chemical reactivity. Many reactions are never explored because the laws that we learn about chemical reactivity discourage us from attempting certain reactions. With all of this in mind, recent research in several groups has looked at new approaches to new reaction discovery. I am going to highlight three of these new approaches in detail.

4.5.1 High-throughput reaction discovery

The Bellomo group at the University of Pennsylvania recently published their work toward the synthesis of HIV integrase inhibitors.[18] In collaboration with scientists at Merck, the Bellomo group targeted a class of molecules with important biological activities. The class of molecules, pyrimidinones, has been shown to inhibit HIV integrase through an unknown mechanism. As the authors point out in their reports, previous attempts to synthesize this class of molecules resulted in typically low overall yields (15- 20%). This is because the reaction conditions require relatively high temperatures (140 ºC) and therefore often lead to decomposition of the desired products.



Scheme 4.17 Discovery of conditions necessary to provide pyrimidone framework

The Bellomo group wanted to investigate the use of a catalyst to improve upon this reaction. They then developed a high-throughput screening platform that would allow them to look at a variety of catalysts under different reaction conditions in order to quickly identify the best catalyst system for the reaction.

The authors first screened a variety of temperature conditions for this reaction and found that the lowest temperature they could use to afford conversion to the desired product was 60 ºC. With this optimized temperature they screened 95 different catalysts in 5 different solvents for a grand total of 475 reactions. In order to complete this process in a high throughput manner a 96 well plate with 250 μL wells was spiked with 20-50% of the catalysts. The substrate, as well as an internal standard, was dissolved in one of five different solvents with a final concentration of 0.125 M. An aliquot of 20 μL of each solution was put into the wells within the plate. The reactions were heated to 60 ºC for 4 hours and then evaluated by HPLC MISER analysis. This allowed the evalutation of 475 reactions by a single scientist in one day.

	1	2	3	4	5	6	7
A	control	LiCl	NaF	NaBr	NaI	NaCN	NaTFA
B	Na_2SO_4	KCl	CsF	CsCl	$V(acac)_2$	$CrCl_2$	$MnCl_2$
C	$Co(acac)_3$	NiF_2	$NiBr_2$ BME	$Ni(acac)_2$	$(PPh_3)_2$ $NiCl_2$	CuCl	$CuCl_2$
D	Mo(acac)	$Pd(OAC)_2$	$PdCl_2dppf$ DCM	$Pd_2(dba)_3$	Ag_2O	AgOTf	AuCl
E	Pd	$Mg(OTf)_2$	$Ca(OTf)_2$	$Sc(OTf)_3$	$Zn(OTf)_2$	$Ga(OTf)_3$	$Y(OTf)_3$

Figure 4.3 Systematic screening of reaction conditions

Once the best catalyst was identified, the authors were able to optimize this reaction. The initial screen identified the best catalyst conditions for the reaction to be dibromo(1, 10-phenanthroline)/ Cu^{II} in 1,4-dioxane (74% assay yield- relative to 50% assay yield with no catalyst). They found that the catalysts loading could be reduced from 50 mol % to 5 mol % without any loss in the overall yield of the reaction. In addition, 1,4-dioxane was replaced with 2-MeTHF and this afforded 84% AY. Finally, the temperature was reduced to 40 ºC to give the optimized yield of 90%.

This approach would allow for the screening of hundreds of diverse reaction conditions in a relatively short amount of time. It should be noted here that this process has a couple important limitations. The first is that only one or two variables can be explored in this system. The authors evaluated different catalysts and different solvent systems. The ideal temperature was identified prior to the analysis. If the authors chose, they could look at a more diverse set of variables but that would increase the complexity of the experiment and therefore increase the time needed for the analysis. In addition, perhaps the greatest limitation to this process is that

the authors know before hand the reaction they are screening for. The known synthesis of pyrimidinone heterocycles from a typical starting material was simply optimized in this process. This process would not be applicable to the discovery of unexpected reactivity.

4.5.2 "Accelerated Serendipity"

David MacMillan's group at Princeton University has also initiated a program aimed at the discovery of new catalyzed reactivity.[19] Their approach, which they refer to as "Accelerated Serendipity", screens multiple reaction variables against one another with the idea that simply by screening a large number of possible conditions, one will find useful conditions in some reactions. What sets the MacMillan lab apart from the work in the Bellomo group is that the MacMillan group is not starting with a bias. Rather than screening a variety of conditions for a specific reaction, they are looking for any type of reactivity.

The method employed by the MacMillan group involves reacting different substrates in combination with various catalysis conditions and analyzing the reactions via GC-MS and using National Institute of Standards and Technology (NIST) mass spectral library software. This allowed them to analyze roughly 1000 reactions each day. Importantly, the substrates that were chosen by the researchers were otherwise benign to each other. In other words, one wouldn't expect any reactivity between the substrates under normal conditions. This would help to eliminate the emergence of known conversions and instead increase the likelihood that novel reactivity will be observed.

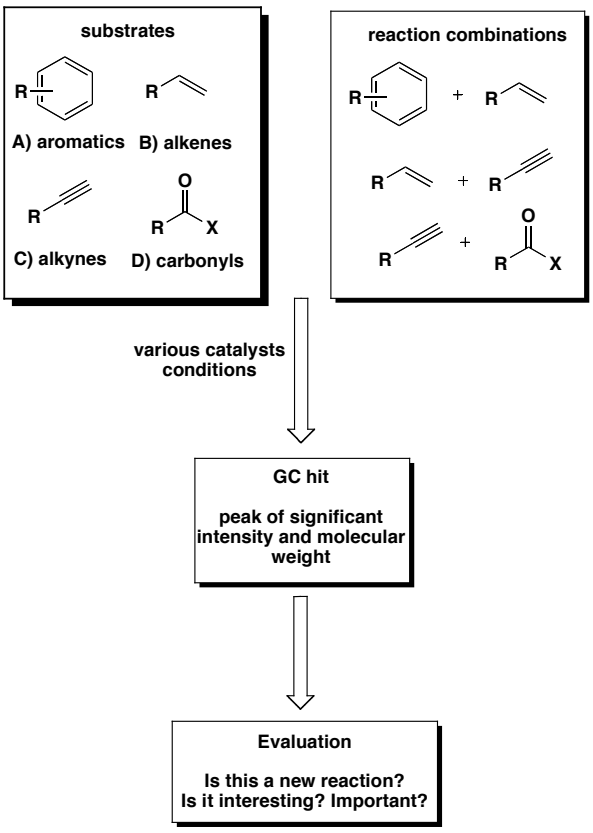
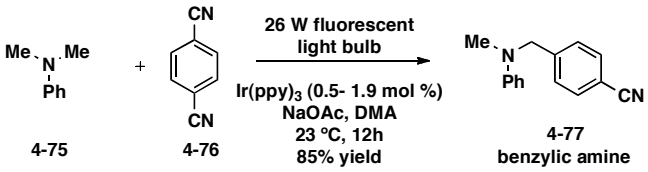


Figure 4.4 "Accelerated Serendipity" discovery platform

Initial "hits" in this assay resulted from a significant new peak with an intensity and molecular weight that was meaningful. The reaction was then evaluated to determine if the reactivity was known or if it was a new reaction. If the reaction was new, was it interesting and worth pursuing? Using this approach, the MacMillan lab discovered a novel photoredox C-H arylation reaction (**Scheme 4.18**).



Scheme 4.18 Photoredox synthesis of benzylic amines

In this process by the MacMillan group it is still important to discuss some limitations. The MacMillan group has access to state of the art equipment and software through collaborations with Merck. Unfortunately, this level of high- throughput screening would be difficult in the typical lab.

4.5.3 Two–dimensional reaction screening

Finally we come to the work by John Hartwig and co-workers.[20] The Hartwig group has been at the front of organometallic catalyzed reactions for several years. Similarly to the MacMillan group, the Hartwig group notes that chance or luck can often lead to exciting discoveries in synthesis. By combining reagents in a vast array of conditions the probability that one will discover a new, exciting reaction increases.

As Hartwig points out, there are some limitations to earlier approaches that do not make them amenable to the average laboratory. In the MacMillan group, the use of expensive equipment precludes this experimentation from the typical synthetic laboratory. In other groups seeking high- throughput analysis, colorimetric tags or the attachment to DNA fragments again preclude the average laboratory. Therefore, they sought to develop a method of multidimensional analysis of catalytic reactions with a method that uses standard laboratory equipment and can be carried forward in useful syntheses.

The initial experiment screened 17 organic reactants each of which contained only a single functional group. If two substrates reacted the product would have a mass greater than any individual substrate and this provided an easy screen for reactivity. In fact, the masses of potential products could be preemptively calculated. The authors combined the masses of the individual substrates and subtracted common small molecule masses and leaving groups (H_2O, NH_3, Cl, etc).

4.6 Conclusion

Biocatalysis is an important tool in organic chemistry as biocatalytic transformations provide several benefits over traditional synthetic methods. For example, biocatalyzed reactions tend to be regio– and stereo specific while using relatively benign reagents and solvents and moderate temperatures. This allows the reactions to be performed on large scales as well. Despite the obvious benefits to biocatalytic processes, the field lacks a method of screening for new reactivity.

In contrast, traditional synthetic chemistry laboratories have focused on new reaction discovery since the beginning of organic chemistry. Recent approaches in this field have expanded on this search for new reactions, and by using the latest technology can now screen large numbers of reaction conditions for new reactivity. However, the approaches that have been applied in this field are not directly amenable to the discovery of new biocatalyzed reactions. The following chapter describes the progress our lab has made toward the development of a new discovery platform to search for novel biocatalyzed organic transformations.

4.7 References

1. Tao, J., Xu, J–H. *Curr. Opinion. In Chem. Bio.* **2009**, *13*, 43–60. Biocatalysis in the development of green pharmaceutical processes.

2. Oliver, A. L., Anderson, B. N., Roddick, F. A., *Adv. Microb. Physiol.* **1999**, *41*, 1–45. Factors affecting the production of L–phenylacetylcarbinol by yeast: a case study.1.

3. Clouthier, C. M., Pelletier, J. N., *Chem. Soc. Rev.*, **2012**, *41*, 1585-1605. Expanding the organic toolbox: a guide to integrating biocatalysis in synthesis.

4. Kille, S., Zilly, F.E., Aceredo, J.P., Reetz, M. T., *Nature Chemistry*, **2011**, *3*, 738-743. Regio- and stereoselectivity of P450- catalyzed hydroxylation of steroids controlled by laboratory evolution.

5. Reed, M. J., Purohit, A., Woo, L. W. L., Potter, B. V. L., *Drugs of the Future*. **2008**, *33*, 597-606. Steroid sulfatase inhibitors for the tropical treatment of skin disorders.

6. Dunn, P. J., Well, A. S., Williams, M. T., Green Chemistry in the Pharmaceutical Industry. Wiley- VCH Verlag GmbH & Co. P 162. **2010.**

7. Whittall, J., Sutton, P. W., Practical Methods for Biocatalysis and Biotransformations 2. 2012. John Wiley and Sons.

8. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., Robins, K., *Nature*, **2012**, *485*, 185-194. Engineering the third wave of biocatalysis.

9. Shing, T. K. M., Lee, C. M., Lo, H. Y., *Tetrahedron Letters*, **2001**, *42*, 8361-8363. Synthesis of the CD ring in taxol from (S)-(+)-carvone.

10. Mino, T, Masuda, S., Nishio, M., Yamashita, M., *Journal of Organic Chemistry*, **1997**, *62*, 2633.

11. Gonzalo, G., Mihovilovic, M. D., Fraaije, M. W., *ChemBioChem*, **2010**, *11*, 2208–2231. Recent developments in the application of Baeyer–Villiger monooxygenases as biocatalysts.

12. Kim, T. K., Hwang, J. Y., Kim, Y. S., Joo, S. H., Change, S. C., Lee, J. S., Takatsuto, S., Kim, S. K., *The Plant Cell*, **2005**, *17*, 2397–2412. Arabidopsis CYP85A2, a Cytochrome P450, mediates the Baeyer–Villiger oxidation of castasterone to brassinolide in brassinosteroid biosynthesis.

13. Cernuchova, P., Mihovilovic, M. D., *Organic & Biomolecular Chemistry*, **2007**, *5*, 1715-1719. Microbial Baeyer- Villiger oxidation of terpenones by recombinant whole- cell biocatalysts- formation of enantiocomplementary regioisomeric lactones.

14. Prarie, R. L., Talalay, P., *Biochemistry*, **1963**, *2*, 203.

15. Kaul, P., Asano, Y., *Microbial Technology*, **2012**, *5*, 18-33. Strategies for discovery and improvement of enzyme function: state of the art and opportunities.

16. Ishige, T., Honda, K., Shimizu, S., *Curr. Opin. Chem. Biol*. **2005**, *9*, 174–180. Whole organism biocatalysis.

17. Kurti, L., Czako, B., Strategic Applications of Named Reactions in Organic Synthesis. 2005. Elsevier Academic Press.

18. Bellomo, A., Celebi-Olcum, N., Bu, Z., Rivera, N., Ruck, R. T., Welch, C. J., Houk, K. N., Dreher, S. D., *Angewandte Chemie, International Edition.* **2012**, *51*, 6912-6915. Rapid Catalyst Identification for the Synthesis of the Pyrimidinone Core of HIV integrase Inhibitors.

19. McNally, A., Prier, C. K., MacMillan, D. W. C., *Science*, **2011**, *334*, 1114-1117. Discovery of an a-Amino C-H Arylation Reaction Using the Strategy of Accelerated Serendipity.

20. Robbins, D. W., Hartwig, J. F., *Science*, **2011**, *333,* 1423-1427. A Simple, Multidimensional Approach to High- Throughput Discovery of Catalytic Reactions.

CHAPTER FIVE

A LABELED SUBSTRATE APPROACH TO THE DISCOVERY OF BIOCATALYTIC REACTIONS

5.1 Development of a discovery platform for new biocatalyzed transformations

It has been well established that biocatalyzed reactions hold an important role in organic synthesis. Presently, synthetic chemists have access to a variety of enzymes for a multitude of oxidation and reduction reactions. However, outside of these known enzymatic transformations, the field lacks a method of screening for *new* biocatalyzed transformations. As a natural products isolation lab, the MacMillan group has created a vast library of diverse bacteria that offer a unique opportunity. This chapter will describe the use of this collection to develop a platform to screen the bacterial collection for enzymatic activity and the optimization of new reactions.

The goal of this project was to develop a medium- throughput method of screening bacterial cultures for the ability to perform useful enzymatic organic transformations. This required the ability to introduce exogenous organic substrates to a culture of bacteria, allow the bacterial enzymes to perform some transformation, recover the organic substrate, and the transformation that occurred through various spectroscopy techniques (**Figure 5.1**).

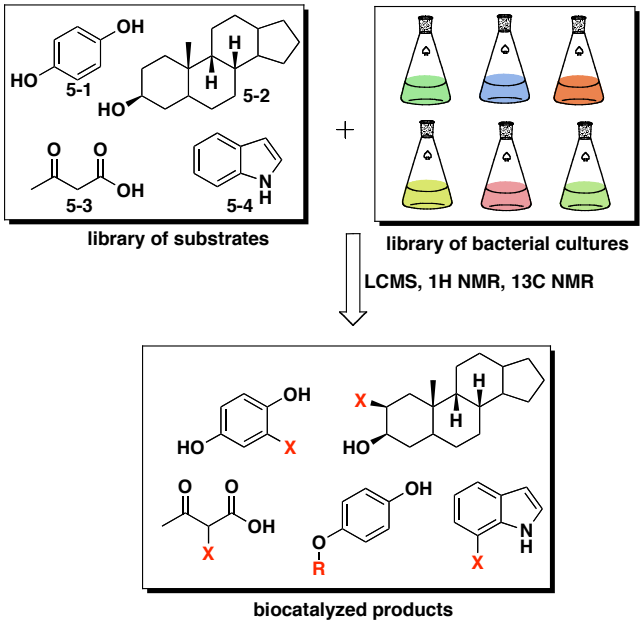


Figure 5.1 Medium–throughput discovery platform for discovery of biocatalyzed organic transformations

Importantly, this discovery platform requires the use of whole–cell bacterial cultures. The use of whole–cell cultures rather than purified enzymes allows this process to be an un–biased screening platform. It is not necessary to identify the active enzyme in the discovery of new reactions. Also, the process of enzyme purification is challenging. Information about the enzyme is necessary to purify it. In addition, many enzymatic reactions require cofactors such as metal salts, $NADH^{+}$, H_2O_2, etc. By utilizing whole–cell cultures, the cofactors can be recycled by the organism so that they are not limiting factors in the reaction.

5.1.1 Proof–of–principle experiment

The first step in developing this process was to show that a culture of bacteria could transform an exogenous substrate through a specific reaction, and that this conversion could be

easily monitored. Early experiments set the groundwork for the development of a discovery platform. Starting with a strain of bacteria (SNA017) from the lab collection, a regioselective halogenation reaction was observed. Using two wells of a 24–well plate, two experiments were performed. The first well was filled with media and no bacterial cells. The second well was filled with SNA017 culture and 0.01 % H_2O_2, a common cofactor. To each well was added purified resistomycin, a bacterial natural product polyketide. In the control well containing only resistomycin and media, no conversion of resistomycin was observed. However, in the well containing bacteria and the H_2O_2 cofactor, approximately 25% conversion to 8-chloro resistomycin was observed by LCMS (**Figure 5.2**).

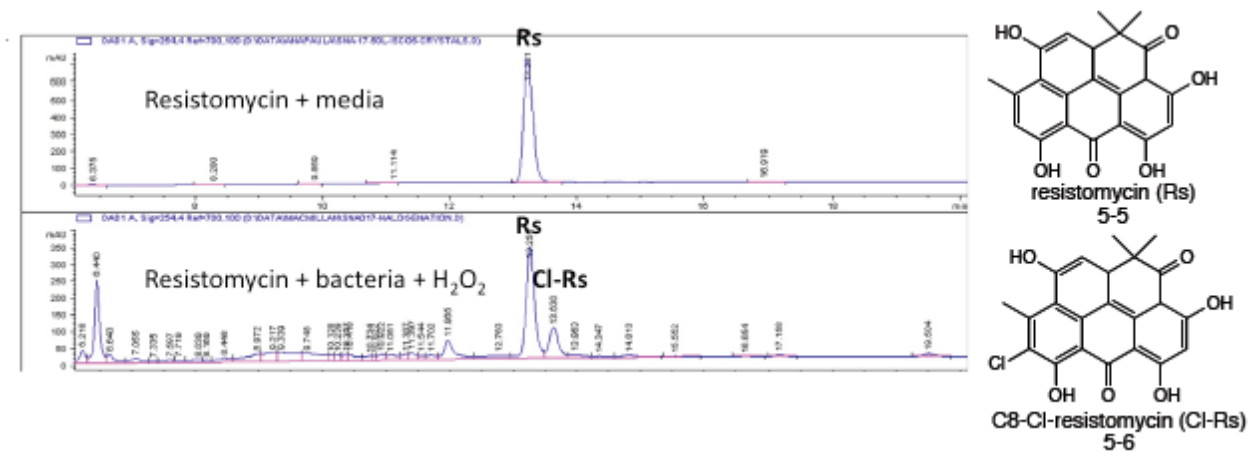


Figure 5.2 Production of Cl-resistomycin through biocatalysis

As a proof–of–principle, this experiment showed that a substrate could be added exogenously to a culture of bacteria and in this reaction, a halogenation reaction was observed via LCMS. However, the goal of this project was to develop a discovery platform without bias for a specific reaction or a particular substrate. In this initial experiment, a known natural

product was added to its producing bacteria and a known halogenation was observed. To develop a successful screening platform it was necessary to show that an un–natural substrate could be added exogenously to a bacterial culture and the reaction could be screened un–biased for new reactions.

5.1.2 Potential complications from the use of whole–cells

Additionally, while the use of whole- cell cultures was important for the success of this experiment, it also produced further hurdles. Many organic substrates and the solvents in which they are soluble can be toxic to bacteria at certain concentrations. To ensure that the substrates were not toxic at the desired concentration, disk assays were performed to monitor the toxicity of the chosen substrates. **Figure 5.3** shows the disc assay for SNA017 when exposed to hydrocinnamic acid. In this assay, varying concentrations of the substrate were exposed to a plate of SNA017 bacteria. At the highest concentration, no toxicity was observed. As the subsequent experiments were performed at lower concentration, there was no concern of toxicity.

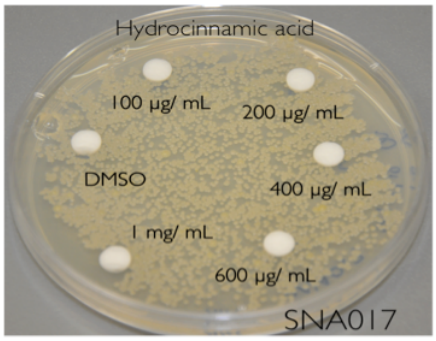


Figure 5.3 Disk assay of hydrocinnamic acid and SNA017

5.1.3 Early detection of substrate oxidation in hydrocinnamic acid

At this time the best approach to analysis of the reactions was via LCMS and ^{1}H NMR as these two approaches were the most sensitive. After determining that hydrocinnamic acid was not toxic to the bacteria at the desired concentration via disk assay, ten milligrams of hydrocinnamic acid was added to a 50 mL culture of SNA017. The reaction progress was monitored at 4, 7, 10 and 17 days. Through ^{1}H NMR, the presence of a new compound was observed (**Figure 5.4**). This compound was identified as ß-oxo hydrocinnamic acid. The identification of a new product was exciting, though the typical oxidation product was not worth follow–up. Although in this example it was possible to identify the presence of a new product, it was quickly realized that the use of whole–cells would cause additional complications.

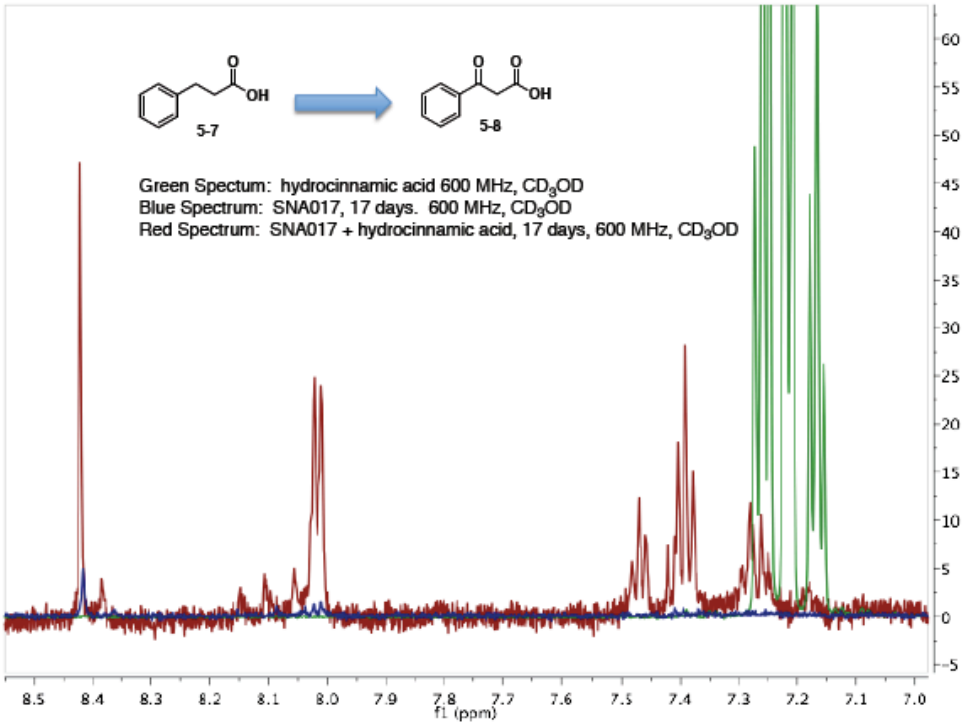


Figure 5.4 Production of ß-oxohydrocinnamic acid

5.1.4 Reducing noise level in analyses

Although the level of noise in the above example was minimal, with other substrate/ culture combinations, the noise in the LCMS chromatogram and ^{1}H NMR made it impossible to detect meaningful changes to the substrate (**Figure 5.5**).

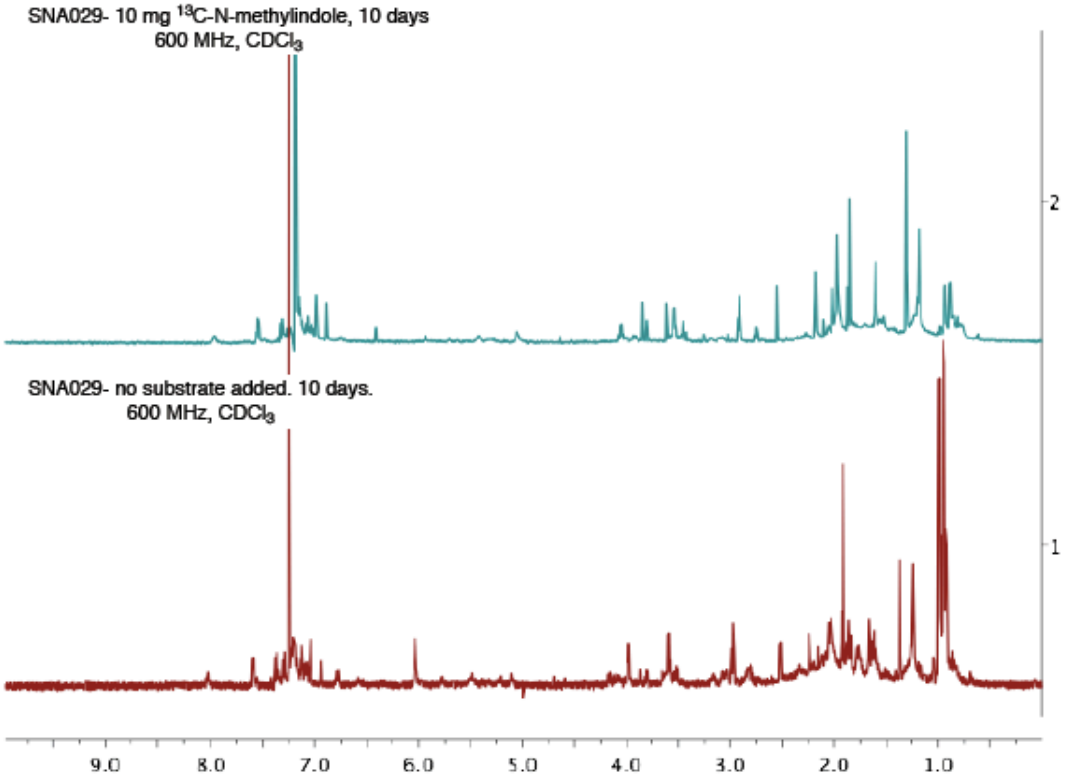
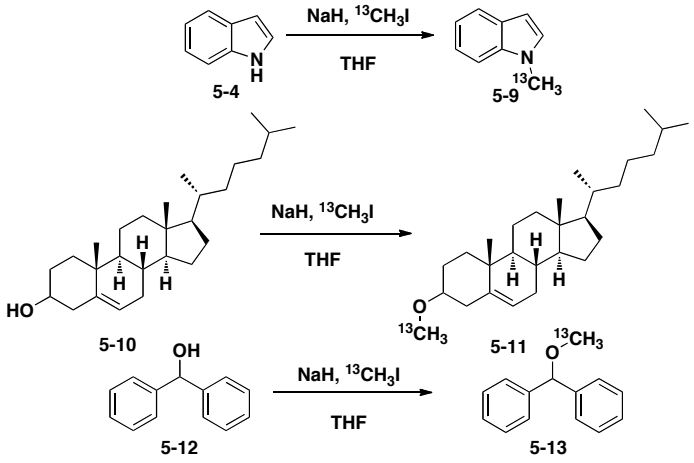


Figure 5.5 Noise in ^{1}H NMR interferes with the analysis

A significant level of noise was observed in these analysis because the bacteria used in these experiments are natural product–producing. The presence of secondary metabolites clouds the spectra and makes it difficult to distinguish between analogs of the exogenous substrate and naturally occurring metabolites. There were two ways in which this was restricted. First, each

experiment was performed with a control to identify and exclude the natural secondary metabolite production. Also, more importantly, an isotope label was added to each substrate as a way of following that substrate through the reaction. The label that was selected was a ^{13}C label as this allowed the reactions to be monitored via ^{13}C NMR.

NMR spectroscopy is one of the most important tools available to organic chemists, as it provides invaluable information about the 2–dimensional and 3-dimenisional structures of organic molecules of all sizes. This technique describes the electronic environment surrounding an atom and allows chemists to distinguish between atoms based on their surroundings and position in the molecule. It is possible to perform NMR spectroscopy with several nuclei, but the most commonly used nucleus is ^{1}H.[1] This is because ^{1}H is the most abundant hydrogen isotope and also a very sensitive nucleus in NMR. Another commonly used nucleus in NMR is ^{13}C. However, this isotope is inherently less sensitive than ^{1}H NMR for two reasons. First, the gyromagnetic ratio, the ratio of the magnetic dipole moment to the angular momentum, of ^{13}C is one forth of that for ^{1}H. In addition, ^{13}C makes up only 1.1% of the total abundance of carbon on earth. This means that it is in very low abundance within organic molecules. Notably, it is possible to change the abundance of ^{13}C by incorporating a ^{13}C label in a molecule. This allows one to use ^{13}C NMR as a very sensitive technique in evaluating the molecule or its derivatives. For this analysis, a ^{13}C label was easily incorporated to several molecules via a substitution reaction (**Scheme 5.1**).



Scheme 5.1 ^{13}C labeling of substrates

Figure 5.6 shows the importance of the ^{13}C label. Both spectra are from the same sample in which ^{13}C-O-methylcholesterol was added exogenously to a culture of SNC011. The spectrum on the left is the ^{1}H NMR spectra of this crude mixture in which the peaks of the cholesterol substrate are obscured by the peaks representing naturally produced metabolites. Alternatively, in the spectrum on the right, the only peak observed is of the labeled cholesterol substrate. The natural abundance of ^{13}C in the metabolites and media components is too low to be observed in a ^{13}C NMR spectrum at this concentration. The use of labeling was essential to the success of this project (**Figure 5.6**).

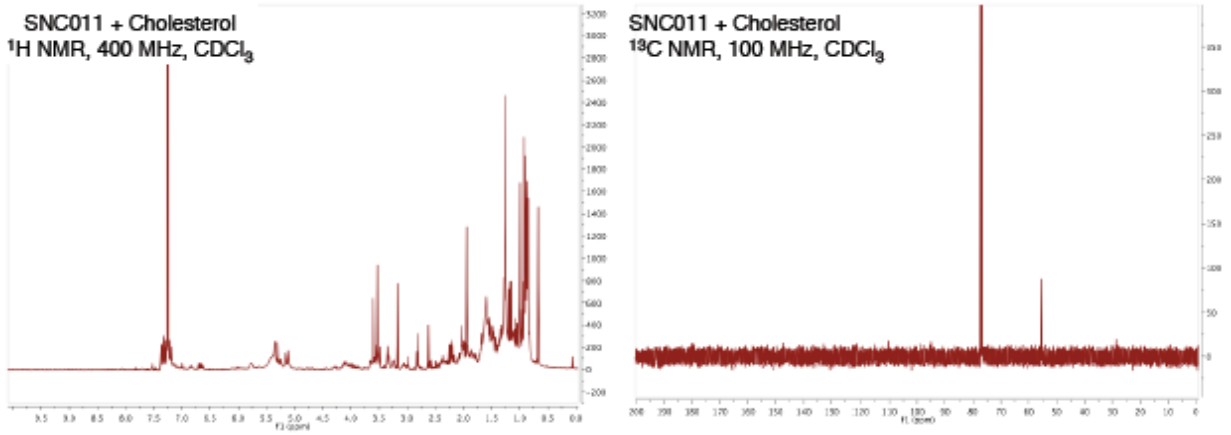


Figure 5.6 Noise reduction with ^{13}C labeling

5.2 Observation of indole oxidation and subsequent optimization

By incorporating a ^{13}C label into the substrates, it was possible to easily follow these substrates from the beginning of the reaction to the end. Starting with five substrates, each compound was added individually to a culture of bacteria. After 4 days, 7 days and 10 days, a 10 mL aliquot of the culture mixture was extracted with ethyl acetate. The organic extract was analyzed via ^{1}H and ^{13}C NMR. Of the 35 initial experiments, one in particular showed promising results. When ^{13}C labeled N-methyl indole was added to a culture of SNB003, complete conversion to a new product was observed. As **Figure 5.7** shows, after four days partial conversion of the original compound to a new product was observed by the up–field shift of the diagnostic ^{13}C label. After ten days the NMR showed complete conversion to the new product.

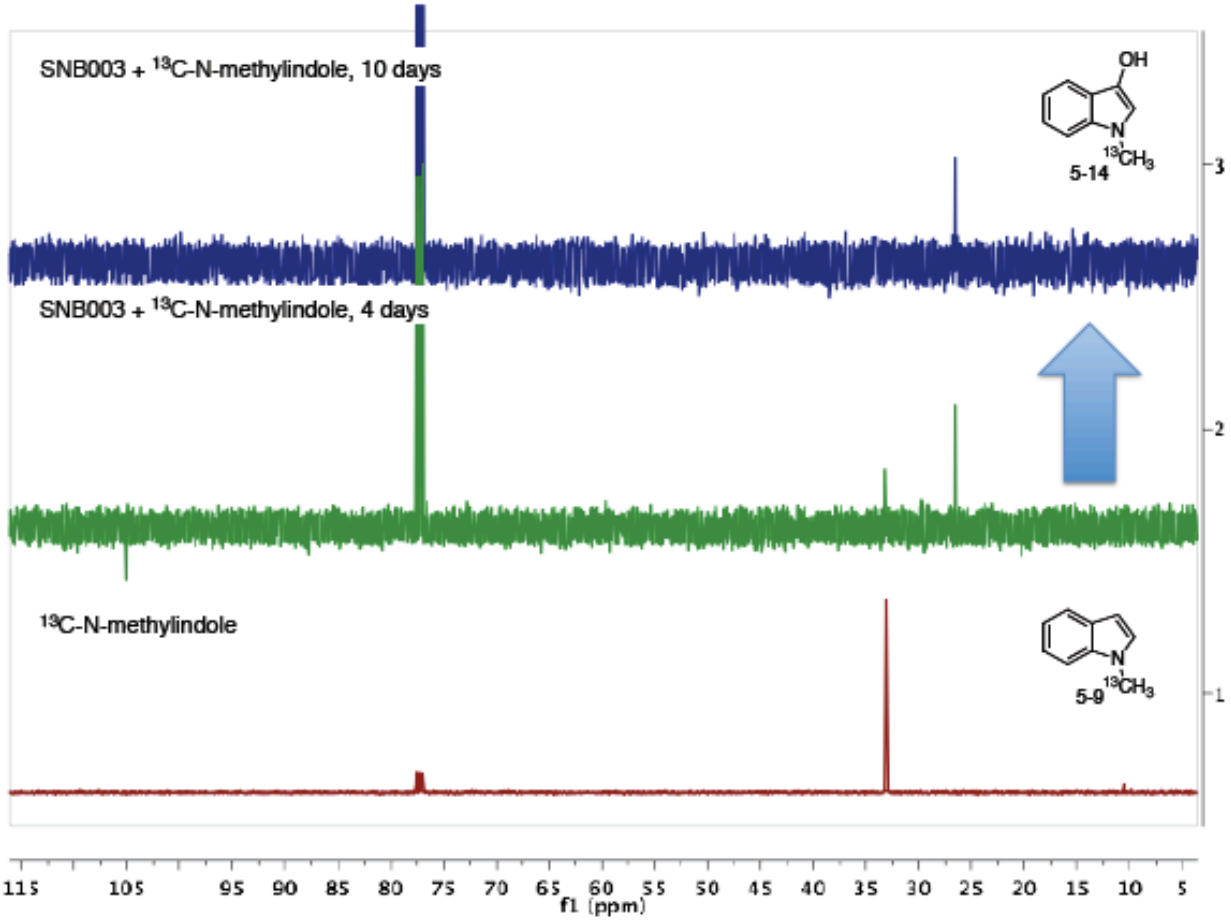
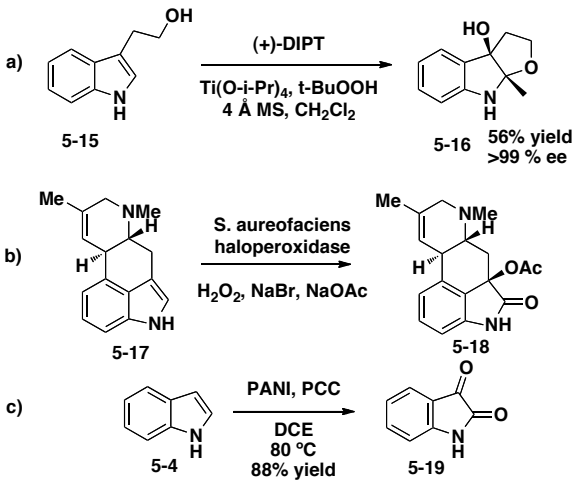


Figure 5.7 Gradual conversion of ^{13}C-N-methylindole to new product

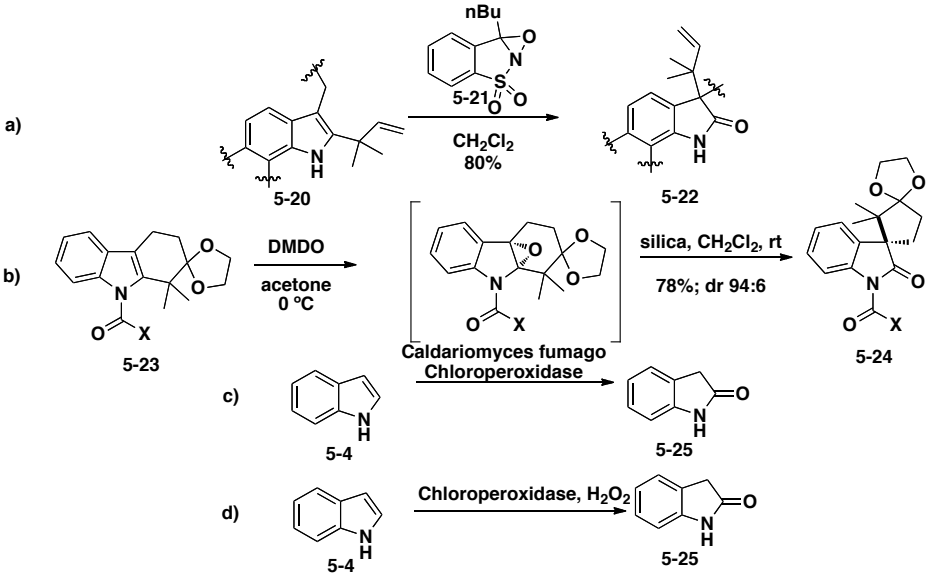
5.2.1 Interesting indole oxidation

The new product was isolated from the crude organic mixture via RP-HPLC on a C–18 column. The structure of this compound was identified as 3-hydroxy-^{13}C-N-methylindole.[2] This result was surprising and exciting for a few reasons. First, oxidation reactions of indole are common. However, the oxidation reaction is notoriously difficult to control. This is illustrated in **Scheme 5.2**, which shows several oxidations of indole all resulting in over–oxidation at the two and three positions.[3-5]



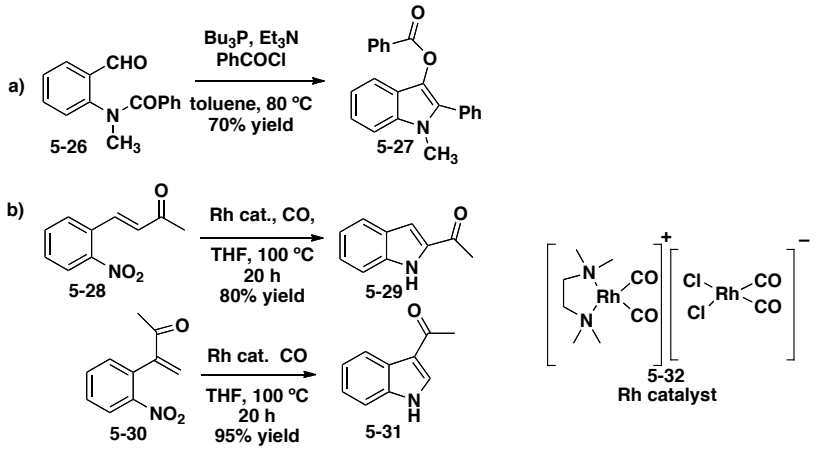
Scheme 5.2 Non-specific indole oxidation

Additionally, when oxidation of indole can be controlled, it is more common to observe the 2-oxindole product (**Scheme 5.3**).[6-9]



Scheme 5.3 Oxidation of indole and derivatives at the 2– position

Because of the difficulty of indole oxidation reactions, the oxidized indole products are often synthesized through a more indirect route. This is illustrated in **Scheme 5.4** where a Wittig reaction[10] and reductive N–heterocyclization, [11] respectively, are employed to afford the desired oxidized indole derivatives.



Scheme 5.4 Synthesis of oxidized indole derivatives

5.2.2 Unexpected indole oxidation with scale–up

To follow up on this initial result, the reaction was scaled up from 50 mL culture in a 125 mL flask with 10 mg substrate to 1.0 L culture in a 2.8 L Fernbach flask with 200 mg of substrate. Again, the reaction progress was monitored via ^{1}H and ^{13}C NMR and again complete conversion to a new single product was observed. However, upon purification of the new product it was discovered that the oxidation reaction had provided the more common indole oxidation product, 2-oxindole. In the scaled up reaction, the only difference was in the type of flask that was used for the experiment, a 125 mL Erlenmeyer flask versus a 2.8 L Fernbach flask. The Fernbach flask is specially designed to allow for greater aeration of the culture. It was

hypothesized that a difference in dissolved oxygen concentration may have been impacting the reaction. When the dissolved oxygen concentration of cultures in each flask was measured, there was a significant difference in the dissolved oxygen concentration of each with the 125 mL flask culture containing a dO_2 concentration of 56% and the 2.8 L flask culture containing a dO_2 concentration of 85% (both measured relative to atmospheric concentration).

5.2.2.1 Evaluating dissolved oxygen levels under different conditions

While it was straight–forward to show that the oxygen concentrations between the cultures of the two flasks differed significantly, it was necessary to show that this was the cause of the differing results. To test this hypothesis, it was necessary to perform the experiment at different dissolved oxygen levels and monitor the production of the oxidation products. We hypothesized that we could alter the oxygen concentration by changing the volume of media in the flask (**Figure 5.8**). This allowed us use the same style of flask between the experiments, but change the dO_2 concentration.

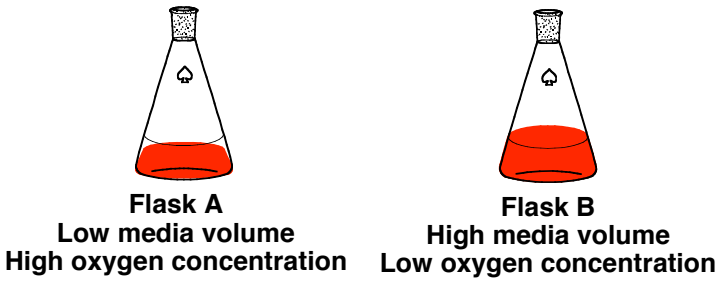


Figure 5.8 Volumes of media varying with dO_2 concentration

5.2.2.2 Evaluating effect of dO_2 concentration on indole oxidation

In an initial experiment, three different flask sizes were used: 125 mL Erlenmeyer flasks, 250 mL Erlenmeyer flasks, and 500 mL Erlenmeyer flasks. The media volume was then varied to allow for differing dissolved oxygen concentrations (**Table 5.1**). Using an Oakton Hand–held Dissolved Oxygen Meter with RS 232C, the relative dO_2 concentrations of each culture was measured after four days of growth. The dO_2 concentration was measured relative to air dO_2 concentration, which was standardized to 100%. The dO_2 concentration was again read after seven days and finally after 11 days.

Although the data showed a clear correlation between the flask size to media volume ratio and dO_2 concentration, inherent flaws complicated the experiment. The small flasks (125 mL) were difficult to measure dO_2 concentration using the Oakton instrument and this caused the measurements to have low precision. In addition, by introducing the probe to the cultures resulted in contamination. We were unable to monitor dO_2 concentration throughout the course of the reaction and determine product concentration after several days because the culture would become contaminated. In order to directly correlate the differing dO_2 to product production, a new experiment was developed.

Table 5.1. Flask size to media volume and the influence on dO_2 concentration

				Air conc. = 100.0% T = 30.1 ºC	Air conc. = 99.7% T = 27.0 ºC	Air conc. = 99.8% T = 26.5 ºC
flask	**media volume**	**flask : media ratio**	**mass of substrate**	**% dO_2 4 days**	**% dO_2 7 days**	**% dO_2 11 days**
125 mL	**25 mL**	**0.20**	**10 mg**	**94.6**	**99.3**	**92.6**
125 mL	**50 mL**	**0.40**	**20 mg**	**101.5**	**98.7**	**96.8**
125 mL	**75 mL**	**0.60**	**30 mg**	**78.2**	**96.1**	**95.4**
250 mL	**50 mL**	**0.20**	**20 mg**	**94.6**	44.5	**77.2**
250 mL	**100 mL**	**0.40**	**40 mg**	**31.3**	**28.7**	**61.7**
250 mL	**150 mL**	**0.60**	**60 mg**	**40.7**	**17.9**	**44.7**
500 mL	**100 mL**	**0.20**	**40 mg**	**43.7**	**51.4**	**88.5**
500 mL	**200 mL**	**0.40**	**80 mg**	**16.4**	**44.9**	**67.6**
500 mL	**300 mL**	**0.60**	**120 mg**	**10.1**	**9.6**	**53.4**

From the previous experiment it was confirmed that simply altering the ratio of flask size to media volume was sufficient to greatly impact the dO_2 concentration within the bacterial cultures. Using an intermediate flask size (500 mL) and three media volumes (100 mL, 200 mL, and 300 mL), a total of 15 cultures were started. To each culture was added ^{13}C-N-methylindole (40 mg/ 100 mL). At various time points (two days, four days, seven days, 10 days, 14 days), the dO_2 concentration was measured in one flask at each volume. Then each culture was extracted with ethyl acetate and analyzed via LCMS and NMR. From this the % conversion to each indole oxidation product relative to starting material was determined (**Table 5.2**).

Table 5.2. Role of % dO_2 on conversion of indole to 2-oxindole or 3-hydroxyindole under whole-cell biotransformation conditions.

Entry	Reaction Time (days)	% 5-9	% 5-33	% Starting Material	Culture Volume (mL)	dO_2 Conc. Relative to Air at 100.0 %
1	14	0	100	0	100	91
2	6	0	100	0	100	89
3	10	0	100	0	100	81
4	4	100	0	0	200	77
5	2	63	0	0	100	73
6	2	100	0	0	200	67
7	4	70	0	0	100	61
8	10	80	0	20	200	61
9	14	75	25	25	200	58
10	10	100	0	0	300	58
11	6	60	40	40	200	50
12	14	100	0	0	300	46
13	6	40	0	60	300	44
14	4	40	0	60	300	39
15	2	30	0	70	300	21

In graphing this data, a clear correlation between dO_2 concentration and production of 2-oxindole vs 3-hydroxyindole is observed. The data show that at low dO_2 concentration conversion is poor and a significant amount of starting material is recovered. Alternatively, at high dO_2 concentration, the conversion to 2–oxindole dominates. This data showed that the ideal concentration of dO_2 is around 60% and this is when conversion to the desired 3–hydroxyindole is favored (**Figure 5.9**).

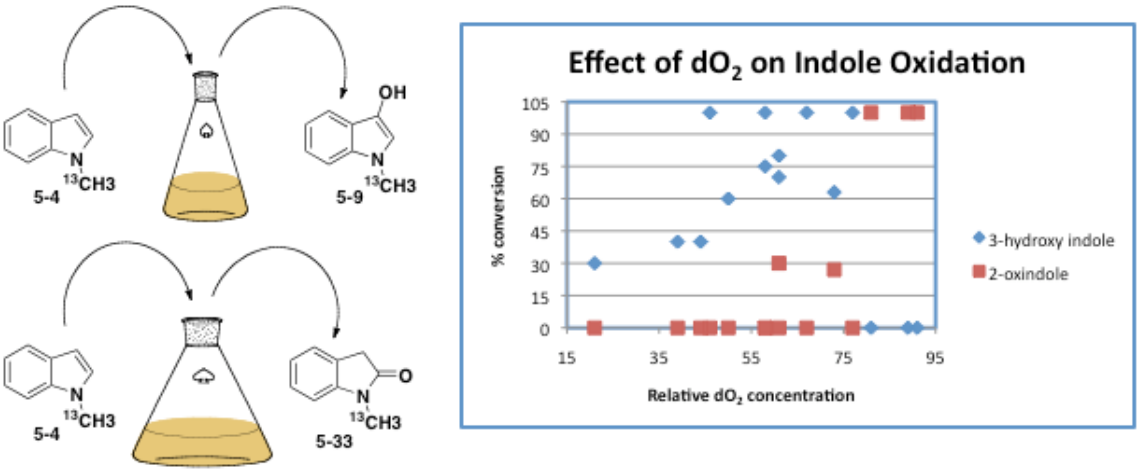


Figure 5.9 Effect of dO_2 concentration on indole oxidation

With optimized conditions in hand, the biocatalyzed indole oxidation was performed again with the idea that under these ideal conditions we could improve upon the yield. However, unfortunately, the highest isolated yield of 3–hydroxyindole was 35%. Interestingly, no side products or starting material were recovered to account for the low mass recovery. We hypothesized that the loss of material was due to incorporation of the substrate by the organism. This possibility was another complication that could arise from the use of whole–cell organisms, and it led us to think of an approach that would eliminate the need for whole–cell organisms while not introducing any bias as discussed earlier.

5.2.3 Enzyme preparation to control indole oxidation

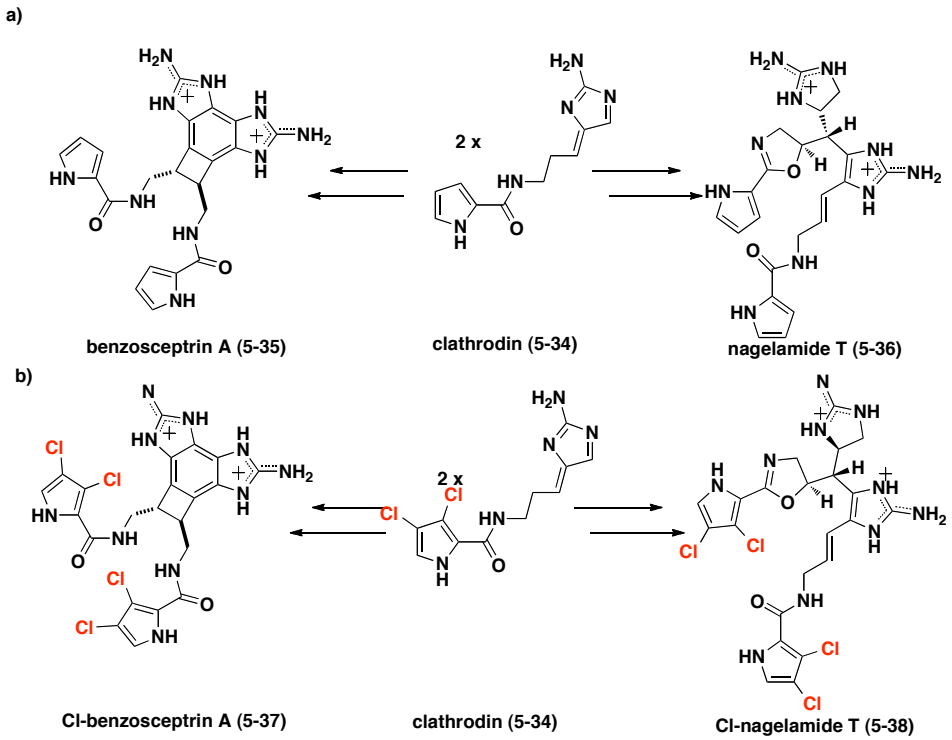
Around this time Dr. Molinski gave a talk at UT Southwestern about work his group had been doing at Scripps Research Institute in La Jolla, California. The Molinski group was interested in a group of natural products, which were each dimers of another natural product, clathrodin (**5-34**). Each of these dimer molecules was produced by marine sponges such as

agelas conifera (**Scheme 5.5a**). The Molinski group was interested in producing un–natural derivatives of these compounds and hypothesized that they could use the organisms' own enzymes to perform the synthesis (**Scheme 5.5b**)

In their initial approach, they took halogenated versions of clathrodin (**5-34**) and added the compound to the environment surrounding the living sponge. They hoped the enzymes within the sponge would perform the same dimerization likely performed in the natural product synthesis, and provide the halogenated dimers. However, they saw no conversion in their initial experiments. They then hypothesized that the complication was that the substrates were never even accessing the enzymes because they could not enter into the cells easily. They chose to prepare an enzyme preparation in which the cells were lysed and the enzymes were indiscriminately removed and used as a crude catalyst for the reaction. With the enzyme preparation they saw conversion of the halogenated clathrodin to the desired dimeric products.

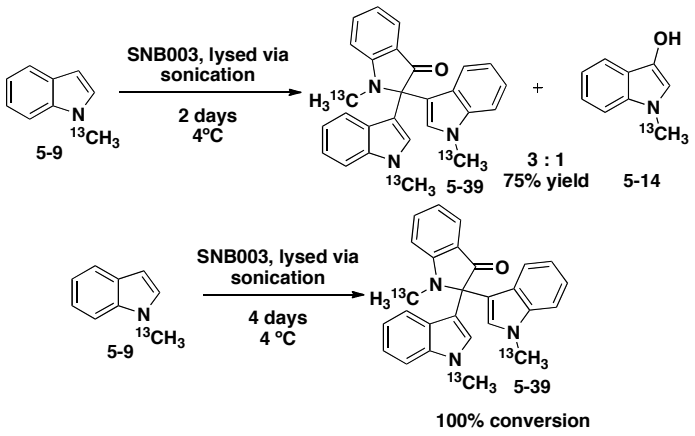
For our purposes, we hoped that by utilizing an enzyme preparation approach we could still access the microbial enzymes without bias and by not using whole–cell catalysis we would not lose material to incorporation by the organism. Many techniques are available to lyse bacterial cells but their success is contingent upon the robustness of the cells. In our initial efforts we utilized a variety of lysis buffers. The cells were shaken in the buffer for several hours. The solutions were centrifuged to force larger cellular particles to settle and the lysate was transferred to sterile flasks where substrate was added. The reactions were monitored over the course of several days but no conversion was observed. In another approach, an Emulsi–Flex–C5 french press was used to lyse the cells. This approach was problematic as the culture had a tendency to

form large clumps and this was not compatible with the press. The third approach utilized sonication to burst the cells.



Scheme 5.5 Clathrodin dimerization to benzosceptrin A and nagelamide T

After lysing the cells via sonication, the culture was centrifuged to remove larger cellular components and the supernatant was transferred to sterile flasks. Again substrate was added, and good conversion of the substrate to products was observed. However, the results were surprising. With the enzyme prep, two major products were observed. The first product was 3–hydroxyindole, but the major product that was recovered was an indole trimer derived through indole oxidation (**Scheme 5.6**). We hypothesized that this trimer synthesis was the result of over–oxidation of indole. By allowing the reaction to continue longer we could see complete conversion to the indole trimer supporting this hypothesis.



Scheme 5.6 Enzyme preparation catalyzed oxidation of indole

5.3 Substrate scope for indole oxidation

To look at the substrate scope of this oxidation reaction, a series of indole derivatives were labeled with the ^{13}C methyl group and allowed to react under the same conditions as the original oxidation reaction. Within this series were electron rich indoles, methoxy and methyl substituted, as well as electron poor indoles, substituted with a nitro group (**Figure 5.10**).

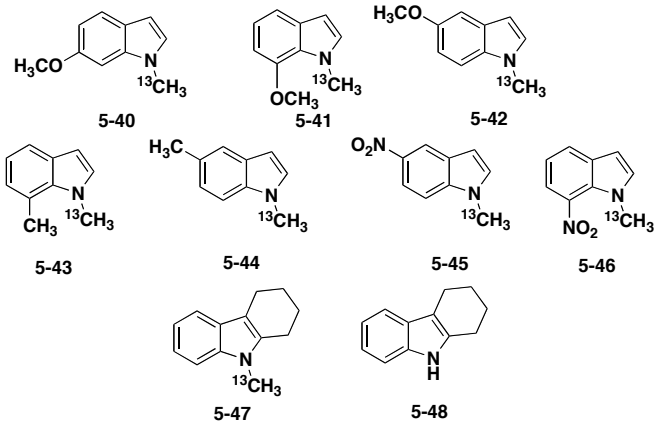


Figure 5.10 Indole derivatives with ^{13}C methyl label

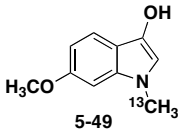
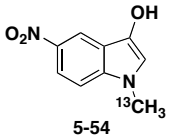
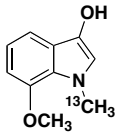
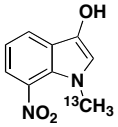
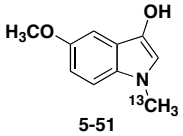
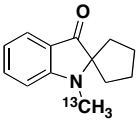
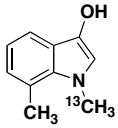
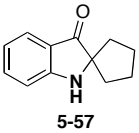
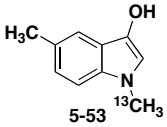
The results of this substrate screen are shown in **Table 5.3**. A trend was quickly observed in this data. Electron rich indoles were generally oxidized to the corresponding 3–hydroxy derivatives while the electron poor indoles were recovered quantitatively as starting material. While electronics may explain this trend in reactivity, the solubility and cell permeability may also be a factor. The nitro–containing indoles are significantly more polar than the other substrates and this could prevent them from crossing into the bacterial cell and therefore prevent oxidation by bacterial enzymes. This hypothesis is supported in part by the quantitative recovery of these substrates. Quantitative mass recovery was not possible with any other substrates, which would suggest that they were consumed in some way by the living organism.

To see if the cell permeability was the problem for the nitro–containing substrates, these substrates were added to an enzyme preparation as described previously. In this experiment, there was no conversion of the substrates observed. Conflicting with our hypothesis, these results suggest that the electron poor substrates were not reactive under these conditions.

Tetrahydrocarbazole (**5-47**) was also an interesting substrate in this screen. This indole derivative is substituted at the 3–position, which is the site of expected oxidation. Unfortunately, the ^{13}C-N-methyltetrahydrocarbozole was not reactive in our reaction conditions. However, the unprotected substrate (**5-48**) was also tested and this molecule provided an interesting rearrangement product as expected with oxidation at the 3–position. This reaction did not go to completion as observed for the other reactive substrates. Starting material was also recovered in

this experiment.

Table 5.3 Substrate scope of indole oxidation

substrate	product	% conversion	substrate	product	% conversion
5-40	 5-49	100	5-45	 5-54	NR
5-41	 5-50	100	5-46	 5-55	NR
5-42	 5-51	NR	5-47	 5-56	NR
5-43	 5-52	100	5-48	 5-57	65
5-44	 5-53	100*			

5.4 Expanding the search for biocatalyzed indole conversions

5.4.1 Exploring other indole reactivity with marine–derived bacteria

After this initial discovery of the selective and tunable indole oxidation reaction, additional bacterial cultures were screened in a search for new biocatalyzed reactivity. To do this, ^{13}C–N–methyl indole was added to the cultures of over 100 bacteria from the MacMillan lab collection. In each case, the bacteria were grown to the stationary phase in 50 mL of media, and then the cultures were centrifuged to collect the bacterial cells. These cells were placed into

fresh A1 media along with 20 mg of the labeled indole substrate. After seven days each culture was extracted with ethyl acetate and the crude material was analyzed via ^{1}H and ^{13}C NMR.

Upon analysis of the reaction mixtures, NMR showed multiple products for each reaction and revealed a very uncontrolled reaction in every experiment. This result was initially surprising as the initial indole oxidation reaction afforded a clean conversion to a single product (**Figure 5.7**). However, upon further analysis this result was actually in line with what is reported in the literature. All of the products identified in these experiments were indole oxidation products and commonly dimeric and trimeric indole compounds were observed including indigo and the indole trimer reported earlier. Enzymatic indole oxidations reported in the literature often result in over–oxidation and specifically in the formation of dimer and trimer molecules. While this result was disappointing, it further added to the intrigue of the indole oxidation catalyzed by strain SNB003. Although the enzyme responsible for this reaction is still unknown, it is clearly significant that it afforded such a clean conversion to 3–hydroxyindole. With the results of this expanded screen, it was apparent that indoles would be easily oxidized in our reaction conditions, and therefore were not ideal substrates for further analysis.

5.4.2 Evaluating complication with ^{13}C based analysis

To further evaluate our method, a group of three compounds were screened in dozens of bacterial cultures from the lab collection. To select the substrates for this broad screen, compounds with specific reactive potential were selected (**Figure 5.11**).

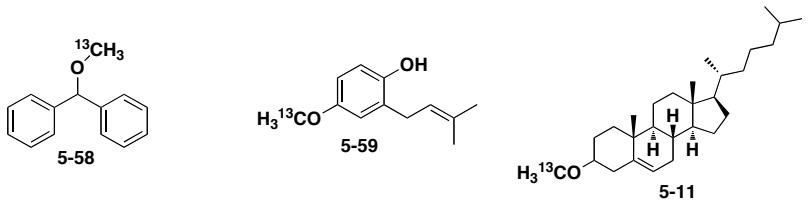
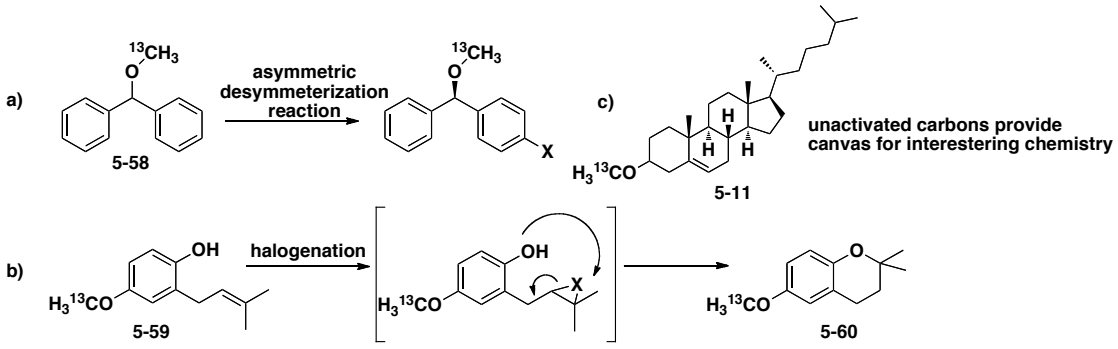


Figure 5.11 ^{13}C labeled organic substrates

The biphenyl compound (**5-40**) was selected because it is prochiral and therefore has the potential of a de–symmeterization reaction (**Scheme 5.7a**). Terpene (**5-41**) was selected as it provided the opportunity to promote a cyclization reaction through initial addition to the alkene (**Scheme 5.7b**). Finally, cholesterol (**5-11**) was selected. This molecule was an attractive substrate as it contains primarily un–activated carbons on which chemistry is traditionally difficult. The discovery of new reactivity with this molecule would be very exciting (**Scheme 5.7**).



Scheme 5.7 Reactivity of selected substrates

Each of the three substrates were labeled with a ^{13}C label and added to cultures of bacteria as described previously. Again after seven days each culture was extracted with ethyl

acetate and analyzed via ^{1}H and ^{13}C NMR. Unfortunately, after evaluation of approximately 30 culture experiments, no conversion of the substrates was observed and the starting material was recovered.

These results were surprising, as it appeared the substrates were completely non–reactive under these reaction conditions. This result also pointed to a potential complication with the developed methodology. To improve the sensitivity of the analysis in these experiments, the ^{13}C label was chosen. This greatly improved the sensitivity of analysis via ^{13}C NMR and this was essential to the discovery of the indole oxidation reaction described previously. However, the ^{13}C label may also be problematic. In order to incorporate the ^{13}C label, it was added to heteroatoms of the desired substrates (**Figure 5.11**). However, the heteroatom can attenuate the electronic environment of the ^{13}C label such that if reactivity occurs on other areas of the molecule, those changes would not cause a chemical shift for the labeled carbon. This in turn actually hurts the sensitivity of this method. It is possible that reactivity had occurred on the chosen substrates but the analysis method was not sensitive enough to detect changes.

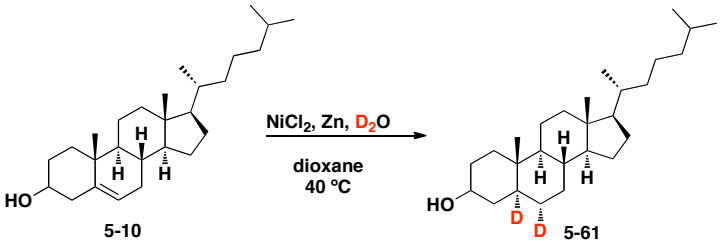
These results led to a search for an alternative method of labeling substrates that would be more sensitive. Labeled carbon atoms are fairly difficult to incorporate directly into the substrates that were evaluated and so a different isotope was desired. Deuterium (^{2}H) was selected as it could be easily incorporated into the main framework of the substrates and it held some of the same advantages as ^{13}C. Specifically, ^{2}H is naturally present in low abundance (roughly .015%),[12] so by adding the isotope selectively to the desired substrate, the abundance of the isotope would be considerable larger in the substrate than in the surrounding environment. By performing ^{2}H NMR, the isotope could be specifically evaluated.

However, ^{2}H NMR also has drawbacks. Specifically, the deuterium signals are much more broad than the corresponding ^{1}H NMR signal. The spin number of the deuterium isotope is 1, which correlates to a longer relaxation time, and therefore line broadening. While ^{1}H NMR has a reference line width of 0.08 Hz, ^{2}H NMR has a reference line width of 1.7 Hz. This also causes ^{2}H NMR to be less sensitive than ^{1}H NMR.[1] However, we chose to move forward with deuterium for this analysis, as it was straightforward to incorporate deuterium into molecules and could allow for selective analysis of our added substrates.

5.5 Exploration for cholesterol transformations

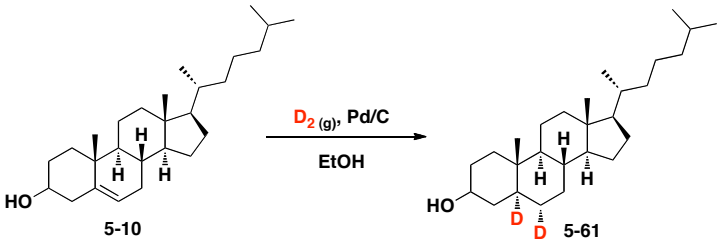
After selecting deuterium as the isotope of choice, a substrate had to be chosen that could be labeled and evaluated in our methodology. Cholesterol was chosen for this analysis. This substrate has a double bond, which could be reduced to afford the saturated molecule and incorporate the deuterium label. Also, through work on the HIF project, sterol type molecules had been shown to be important in an unexpected biological pathway and this increased our interest in this molecule. We hoped to discover ways of synthesizing cholesterol derivatives that could potentially be evaluated as regulated of the hypoxic response pathway.

The deuterium isotope is available in a variety of commercially available reagents. One readily available source is deuterated water (D_2O). Initial efforts to label the cholesterol substrate involved reaction of cholesterol with $NiCl_2$ catalyst and zinc metal with D_2O as the source of deuterium (**Scheme 5.8**).[13] Unfortunately, this reaction was unsuccessful with this substrate and only starting material was recovered from this reaction.



Scheme 5.8 Zinc catalyzed reduction of cholesterol

Another convenient source of deuterium is deuterium gas (D_2). The reduction of cholesterol was performed using palladium on carbon and D_2 gas to give clean conversion to the desired product with high yield (**Scheme 5.9**).[14]



Scheme 5.9 Palladium catalyzed reduction of cholesterol

Unfortunately, the NMR peaks in the spectra of this compound were too broad and this analysis was much less sensitive than we had anticipated. More than 30 mg of the substrate were necessary to obtain a clear NMR after 20 minutes. This required the biocatalysis experiments to be run on a larger scale than originally anticipated and a larger addition of the substrate (80 mg) was required. Nonetheless, this material was then carried forward into the biocatalysis experiments through the same protocol described previously on a scale of 300 mL of culture. Upon analysis of the crude organic material, no change in the ^2H NMR was observed. It was not

clear if the broad signals prevented any shift changes from being observed, or if no biotransformations were occurring with the substrate. These results confirmed that the deuterium label would not be a suitable substitute for the ^{13}C label.

5.6 Conclusion

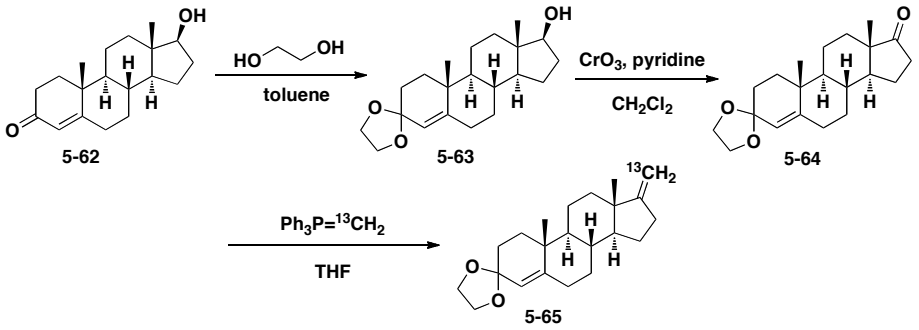
This work has led to the successful development of a discovery platform for the identification of new biocatalyzed reactions by analyzing a diverse collection of marine–derived bacteria. As a proof–of–principle reaction, the selective and controlled oxidation of ^{13}C-N-methylindole was optimized. This reaction can be tuned to afford 3–hydroxyindole, 2–oxindole, or the indole trimer (**5-39**) depending upon the reaction conditions.

Unfortunately, further analysis of indole reactivity revealed uncontrolled oxidation reactions that were not practical for use in organic synthesis. Further evaluation of other substrates revealed the opposite problem in that these substrates showed no reactivity under the reaction conditions. This led us to reevaluate the methodology and the potential complications caused by the labeling approach. Specifically, by adding the ^{13}C label to a heteroatom on the substrates, the label was isolated electronically from the rest of the molecule. As a result, synthetic changes to the molecule may not be reported through spectrometric analysis of the ^{13}C label.

To circumvent this problem, alternative labeling approaches were necessary. Cholesterol was therefore labeled with deuterium and applied to a biocatalysis experiment. While deuterium could be more directly incorporated into the molecule, ^{2}H NMR proved to be a prohibitively insensitive technique and was not successful.

With this groundwork laid, this methodology opens many avenues for further

experimentation. The biggest hurdle in this project is to find the best analysis method of these experiments or more specifically, the best labeling method of the substrates. There are two potential approaches for how to proceed. First, the ^{13}C label was sufficiently sensitive in the analysis but by incorporating it on a heteroatom, its ability to detect changes to the substrate was limited. If this label was more directly incorporated on the molecule this could correct this problem. For a sterol type substrate, the label could be incorporated through a Wittig olefination (**Scheme 5.10**).[15-18] In this example the label would be closer to the main framework of the substrate and would hopefully respond to changes within the molecule more easily.



Scheme 5.10 ^{13}C incorporation on testosterone

A second solution to the labeling conflict would be to use a radioisotope label. ^{14}C is a radioactive isotope of carbon and is present in even lower natural abundance than ^{13}C (one part per trillion) and would therefore be easily distinguishable from the other naturally occurring molecules. Radioisotopes can be detected via autoradiography and can be detected at much lower concentrations than can be seen with NMR.[19] ^{14}C is superior to ^{13}C because it would be detected via TLC rather than NMR. Therefore, the electronic environment surrounding the label would not matter. Instead, the reaction mixtures would be run on a TLC where each individual

molecule would be separated based on polarity. Small changes to a molecule could result in significant polarity changes (**Figure 5.12**).

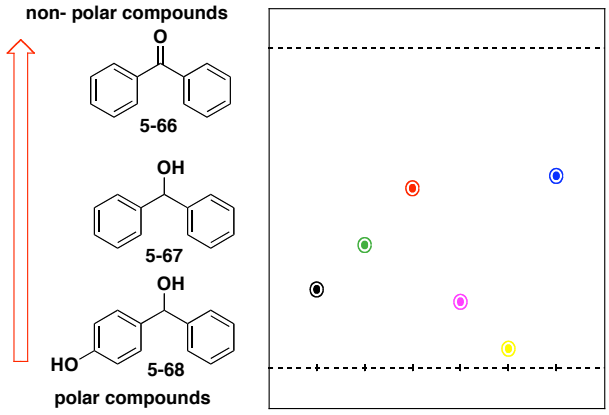


Figure 5.12 Separation of compounds via TLC

Another advantage to the ^{14}C label is that with the increase in sensitivity, this screening platform could be applied on a larger scale and could become a high throughput analysis. However, there are also additional complications that arise with the use of this isotope. Radioactive labels are dangerous and therefore their use requires special training and special lab equipment. Although radioactive labels decompose and therefore have a limited half–life, the half–life of ^{14}C is over 5,000 years and so that will not be a factor in this analysis.

5.7 Experimental

5.7.1 Materials and methods

All reactions were performed under an atmosphere of nitrogen in flame-dried glassware and with dry solvents. Commercially obtained reagents were used as received. Reactions were monitored by thin-layer chromatography on silica gel plates (60 F254 pre-coated plates (0.25 mm). Flash

chromatography was performed using silica gel (particle size 0.032-0.063 mm) purchased from SiliCycle. ^{1}H and ^{13}C NMR spectral data were recorded at 600 MHz or 400 MHz and 100 MHz respectively in $CDCl_3$, and chemical shifts were referenced to the corresponding solvent residual signal. Coupling constants are reported as three bond H-H coupling unless otherwise noted. Low-resolution LC/ESI-MS data were measured using ESI-MS with a reversed phase C_{18} column (Phenomenex Luna, 150 mm x 4.6 mm, 5 mm) at a flow rate of 0.7 mL/ min.

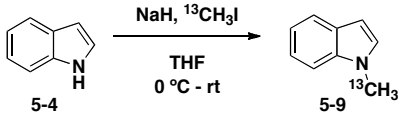
5.7.2 Collection and phylogenetic analysis of strain SNB-003:

The marine-derived bacterium strain SNB-003, was isolated from sediment sample collected from Trinity Bay, Galveston, TX (29° 42.419'N, 94° 49165' W). Bacterial spores were collected via stepwise centrifugation as follows: 2 g of sediment was dried over 24 h in an incubator at 35 °C and the resulting sediment added to 10 mL sH_2O containing 0.05% Tween 20. After a vigorous vortex for 10 min, the sediment was centrifuged at 18000 rpm for 25 min (4 °C) and the resulting spore pellet collected. The resuspended spore pellet (4 mL sH_2O) was plated on an acidified Gauze media, giving rise to individual colonies of SNB-003 after two weeks. Analysis of the 16S rRNA sequence of SNB-003 revealed 98% identity to *Salinospora arenicola.*

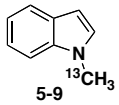
5.7.3 General Procedure for Cultivation of Bacteria and Whole- cell Biotransformation: Bacterium SNB-003 was cultured in a 125 mL Erlenmeyer flask containing 50 mL of a 0.75% seawater-based A1 medium (0.5 g starch, 0.2 g yeast extract, 0.1 g peptone) and shaken at 200 rpm at 33 °C. After 4-14 days when the cells had reached an acceptable density the culture was transferred to sterile 50 mL centrifuge tubes and centrifuged at 5200 RPM for 15 min. The supernatant was decanted off and the remaining cells were transferred to 50 mL sterile media in

125 mL Erlenmeyer flasks. To each flask was added 10 mg of substrate. The reactions were monitored at 4, 7, and 10 days by removing a 10 mL aliquot and extracting with ethyl acetate (3 x 10 mL). After 10 days the remaining culture was extracted with ethyl acetate (3 x 30 mL). The organic phase was conc. *in vacuo* and the products analyzed via ^{1}H, ^{13}C NMR and LCMS. The crude extract was purified via RP HPLC (Phenomenex phenyl-hexyl, 250 x 10.0 mm, 2.5 mL/ min, 5 mm, UC = 254 nm) using a gradient solvent system from 20% CH_3CN to 90% CH_3CN (0.1% FA) over 30 min.

5.7.4 Representative procedure for ^{13}C–N–methylation of indole substrates for biotransformation experiment

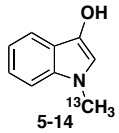


^{13}C labeled substrates were prepared by standard methylation. To a solution of indole (480 mg, 4.10 mmol) in THF (14 mL, 0.3 M) at 0 °C was added sodium hydride (60% in mineral oil, 492 mg, 12.3 mmol). The reaction was warmed to room temperature and allowed to stir for 30 min as a dark blue color developed. After 30 min, the reaction flask was cooled to 0 °C and ^{13}C-methyl iodide (257 µL, 4.10 mmol) was added, drop wise. The reaction was warmed to room temperature and allowed to stir until all starting material was consumed. The reaction was cooled to 0 °C and quenched with saturated NH_4Cl. The product was extracted with ether (3 x 15 mL) and washed with H_2O (45 mL) and sat. NaCl (45 mL). The organic phase was conc. *in vacuo* to give a yellow liquid. The crude material was purified via flash chromatography on silica gel (10 % ethyl acetate in hexane) to give N-^{13}C-methylindole (513 mg , 95% yield) as a clear liquid.Su



5-9

^{1}H NMR ($CDCl_3$, 600 MHz) 7.63 (d, J = 7.9 Hz, 1H), 7.32 (d, J = 8.2 Hz, 1H), 7.22 (dd, J = 7.3, 7.3 Hz, 1H), 7.11 (dd, J = 7.7, 7.7 Hz, 1H), 7.05 (dd, J = 2.6, 2.6 Hz, 1H), 3.79 (d, $^{1}J_{CH}$ = 138 Hz, 3H) ^{13}C NMR ($CDCl_3$, 125 MHz) 136.8, 129.0, 128.6, 121.6, 121.0, 119.4, 109.4, 101.1 LCMS (ESI) calc'd for $[C_8{}^{13}CH_{10}N]^+$([M+H]$^+$): *m/z* 133.17, found 133.2.

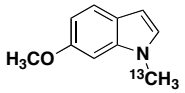


5-14

^{1}H NMR ($CDCl_3$, 600 MHz) 7.43 (d, J = 7.7 Hz, 1H), 7.34 (dd, J = 7.7, 7.7 Hz, 1H), 7.11 (dd, J = 7.2, 7.2 Hz, 1H), 7.04 (s, 1H), 6.81 (d, J = 8.34 Hz, 1H), 5.03 (s, 1H), 3.18 (d $^{1}J_{CH}$ = 140 Hz, 3H) ^{13}C NMR ($CDCl_3$, 100 MHz) 215.7, 130.2, 125.4, 123.4, 111.5, 109.7, 108.7, 108.5, 77.4 LCMS (ESI) calc'd for $[C_8{}^{13}CH_8NO]^+$([M-H]$^-$): *m/z* 147.16, found 147.

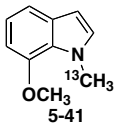
5-33

^{1}H NMR ($CDCl_3$, 600 MHz) 7.27 (dd, J = 7.7 Hz, 1H), 7.22 (d, J = 7.3 Hz, 1H), 7.02 (dd, J = 7.4 Hz, 1H), 6.80 (d, J = 7.8 Hz, 1H), 3.50 (s, 2H), 3.19 (d, J = 139 Hz, 3H). ^{13}C NMR ($CDCl_3$, 125 MHz) 175.3, 145.4, 128.1, 124.7, 124.5, 122.5, 108.3, 36.0, 26.4 LCMS (ESI) calc'd for $[C_8{}^{13}CH_{10}NO]^+$([M+H]$^+$): *m/z* 149.17, found 149.1



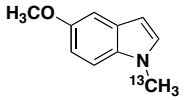
5-40

^{1}H NMR ($CDCl_3$, 400 MHz) 7.47 (d, J = 9.2 Hz, 1H), 6.93 (dd, II = 2.2, 2.2 Hz, 1H), 6.77 (m, 2H), 6.39 (d, J = 3.0 Hz), 3.87 (s, 3H), 3.72 (d, $^1J_{CH}$ = 138 Hz, 3H) ^{13}C NMR ($CDCl_3$, 100 MHz) 156.4, 137.5, 127.9, 121.6, 109.4, 100.9, 92.9, 92.8, 55.9, 33.0 LCMS (ESI) calc'd for $[C_9{}^{13}CH_{11}NO]^+$($[M+H]^+$): *m/z* 163.19, 163.20



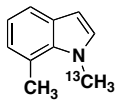
5-41

^{1}H NMR ($CDCl_3$, 400 MHz) 7.17 (d, J = 8.1 Hz, 1H), 6.94 (dd, J = 7.6 Hz, 1H), 6.89 (s, 1H), 6.57 (d, J = 7.6 Hz, 1H), 6.38 (d, J = 3.0 Hz, 1H), 3.90 (s, 3H), 2.34 (d, $^1J_{CH}$ = 139 Hz, 3H). ^{13}C NMR ($CDCl_3$, 100 MHz) 154.1, 129.5, 129.5, 129.0, 112.1, 110.1, 102.7, 100.6, 56.1, 33.2 LCMS (ESI) calc'd for $[C_9{}^{13}CH_{11}NO]^+$($[M+H]^+$): m/z 163.19, 163.20



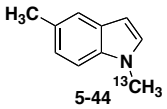
5-42

^{1}H NMR ($CDCl_3$, 400 MHz) 7.15 (d, J = 8.8 Hz, 1H), 6.96 (m, 2H), 6.58 (d, J = 8.7 Hz, 1H), 6.52 (d, J = 8.7 Hz, 1H), 6.38 (bs, 1H), 3.96 (s, 3H), 3.75 (d, J = 138 Hz, 3H) ^{13}C NMR ($CDCl_3$, 100 MHz) 154.1, 129.5, 129.5, 129.0 112.1, 110.1, 102.7, 100.6, 56.1, 33.2 LCMS (ESI) calc'd for $[C_8{}^{13}C_2H_{12}NO]^+$($[M+H]^+$): *m/z* 164.19, found 164.2



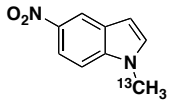
5-43

^{1}H NMR ($CDCl_3$, 400 MHz) 7.44 (d, *J* = 7.9 Hz, 1H), 6.93 (m, 3H), 6.42 (d, *J* = 3.0 Hz, 1H), 4.06 (d, $^1J_{CH}$ = 138 Hz, 3H), 2.77 (s, 3H) ^{13}C NMR ($CDCl_3$, 100 MHz) 130.6, 129.8, 124.3, 121.4, 119.7, 119.3, 101.1, 101.1, 37.0, 20.0 LCMS (ESI) calc'd for $[C_9{}^{13}CH_{10}N]^-$([M-H]$^-$): *m/z* 145.19 found 145.1



5-44

NMR ($CDCl_3$, 600 MHz) 7.39 (s, 1H), 7.20 (d, *J* = 8.3 Hz, 1H), 7.03 (d, *J* = 8.3 Hz, 1H), 6.99 (bs, 1H), 3.75 (d, $^1J_{CH}$ = 138 Hz, 3H), 2.43 (s, 3H) ^{13}C NMR ($CDCl_3$, 125 MHz) 128.9, 121.6, 119.4, 109.3, 104.9, 101.0, 101.0, 32.9 LCMS (ESI) calc'd for $[C_9{}^{13}CH_{10}N]^-$([M-H]$^-$): *m/z* 145.19 found 145.1

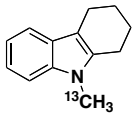


5-45

^{1}H NMR ($CDCl_3$, 600 MHz) 8.57 (d, *J* = 2.1 Hz, 1H), 8.11 (s, 1H), 7.32 (d, *J* = 9.06 Hz, 1H), 7.19 (dd, *J* = 2.1 Hz, 1H), 7.65 (d, *J* = 3.1 Hz, 1H), 3.84 (d, $^1J_{CH}$ = 139.2 Hz, 3H) ^{13}C NMR ($CDCl_3$, 100 MHz) 141.8, 139.6, 132.2, 127.8, 118.4, 117.5, 109.2, 104.0, 33.5 LCMS (ESI) calc'd for $[C8{}^{13}CH_8N_2O_2]^+$([M+H]$^+$): *m/z* 178.06 found 178.1

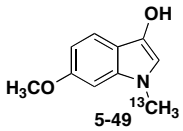
5-46

^{1}H NMR ($CDCl_3$, 600 MHz) 8.15 (d, J = 7.9 Hz, 1H), 7.96 (d, J = 7.7 Hz, 1H), 7.39 (dd, J = 4.1, 4.1 Hz, 1H), 7.20 (dd, J = 8.0, 8.9 Hz, 1H), 6.70 (dd, J = 2.2, 3.2 Hz, 1H), 3.84 (d, $^{1}J_{CH}$ = 140 Hz, 3H) ^{13}C NMR ($CDCl_3$, 100 MHz) 129.4, 123.8, 120.3, 119.6, 113.6, 103.0, 102.0, 101.9, 36.9 LCMS (ESI) calc'd for $[C_8{}^{13}CH_9N_2O_2]^-$([M+H]$^+$): *m/z* 178.16 found 178.2



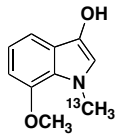
5-47

^{1}H NMR ($CDCl_3$, 600 MHz) 7.45 (d, J = 7.9 Hz, 1H), 7.23 (s, 1H), 7.14 (dd, J = 7.1, 7.1 Hz, 1H), 7.05 (dd, J = 7.6, 7.6 Hz, 1H), 3.60 (d, $^{1}J_{CH}$ = 129 Hz, 3H), 2.71 (m, 4H), 1.94 (m, 2H), 1.84 (m, 2H) ^{13}C NMR ($CDCl_3$, 100 MHz) 136.9, 135.9, 127.3, 120.7, 118.7, 117.9, 109.4, 108.6, 29.1, 23.5, 23.4, 22.3, 21.3 LCMS (ESI) calc'd for $[C_{12}{}^{13}CH_{16}N]^+$([M+H]$^+$): *m/z* 187.26 found 187.1^{1}H



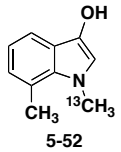
5-49

^{1}H NMR ($CDCl_3$, 400 MHz) 7.32 (d, J = 8.1 Hz, 1H), 6.56 (d, J = 8.8 Hz, 1H), 6.38 (s, 1H), 4.97 (s, 1H), 3.82 (s, 3H), 3.48 (d, $^{1}J_{CH}$ = 140 Hz, 3H), 2.81 (bs, 1H) ^{13}C NMR ($CDCl_3$, 100 MHz) 161.5, 138.8, 127.6, 126.0, 110.0, 106.6, 96.7, 96.6, 55.6, 26.2 LCMS (ESI) calc'd for $[C_9{}^{13}CH_{10}NO_2]^-$([M-H]$^-$): *m/z* 177.18 found 177.2



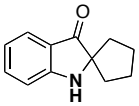
5-50

^{1}H NMR ($CDCl_3$, 400 MHz) 6.97 (dd, *J* = 9.0 Hz, 1H), 6.86 (m, 2H), 5.32 (bs, 1H), 3.94 (s, 3H), 3.76 (d, $^1J_{CH}$ = 140 Hz, 3H), 3.50 (s, 1H). ^{13}C NMR ($CDCl_3$, 100 MHz) 154.0, 136.3, 129.6, 128.8, 122.0, 112.0, 110.1, 100.6, 55.9, 32.9 LCMS (ESI) calc'd for $[C_9{}^{13}C\ H_{10}NO_2]^-$([M-H]$^-$): *m/z* 177.18 found 177.2



5-52

^{1}H NMR ($CDCl_3$, 400 MHz) 7.26 (d, *J* = 6.5 Hz, 1H), 7.16 (s, 1H), 7.05 (d, *J* = 9.7 Hz, 1H), 6.97 (dd, *J* = 6.7, 6.7 Hz, 1H), 4.95 (bs, 1H), 3.44 (d, $^1J_{CH}$ = 134 Hz, 3H), 2.54 (s, 3H) ^{13}C NMR ($CDCl_3$, 100 MHz) 134.5, 128.3, 124.5, 122.5, 120.5, 110.9, 108.3, 90.2, 36.1, 26.4 LCMS (ESI) calc'd for $[C_9{}^{13}CH_{10}NO]^-$([M-H]$^-$): *m/z* 161.19 found 161.2

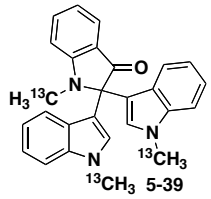


5-57

^{1}H NMR ($CDCl_3$, 400 MHz) 7.61 (d, *J* = 7.2 Hz, 1H), 7.44 (dd, *J* = 7.8, 7.8 Hz, 1H), 6.83 (s, 1H), 6.82 (dd, *J* = 7.0, 7.0 Hz, 1H), 2.08 (m, 2H), 1.97 (m, 2H), 1.83 (m, 2H), 1.71 (m, 2H) ^{13}C NMR ($CDCl_3$, 100 MHz) 137.1, 124.9, 119.0, 112.4, 77.4, 53.7, 38.3, 25.6 LCMS (ESI) calc'd for $[C_{12}H_{12}NO]^-$([M-H]$^-$): *m/z* 186.23 found 186.2

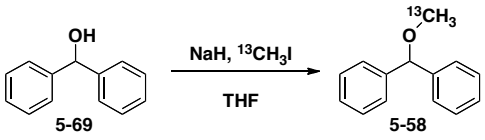
5.7.5 Cell–free enzyme preparation

A 200 mL culture of SNB003 was grown in a 500 mL flask at 33 °C for 10 days. A 50 mL aliquot was removed from the flask and centrifuged at 5200 RPM for 15 min at 4 °C. The supernatant was removed, 50 mL of culture were added and the process repeated two more times. The combined cells were vortexed for 1 min and then placed on ice for 30 min. This solution was sonicated for 5-second intervals and cooled on ice between sonication periods. The cells were sonicated for a combined 5 minutes. The lysed cells were then centrifuged at 5200 RPM for 15 min. The supernatant was transferred to a sterile 125 mL Erlenmeyer flask covered with cheese cloth. To the lysate was added 10 mg of N-[^{13}C]-methylindole. The reaction was monitored at 2, 4 and 7 days by removing a 10 mL aliquot and extracting with ethyl acetate. The organic phase was conc. *in vacuo* and analyzed via ^{1}H and ^{13}C NMR. The samples were then purified via reversed phase HPLC on a Phenomenex phenylhexyl column.

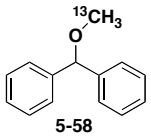


^{1}H NMR ($CDCl_3$, 600 MHz) 7.63 (d, J = 7.2 Hz, 1H), 7.54 (dd, J = 7.7. 7.7 Hz), 7.33 (d, J = 8.8 Hz, 2H), 7.27 (d, J = 7.9 Hz, 2H), 7.17 (dd, J = 8.2, 8.2 Hz, 2H), 6.98 (s, 2H), 6.94 (dd, J = 7.3, 7.3 Hz, 2H), 6.81 (d, J = 9.4 Hz, 1H), 6.73 (dd, J = 8.6 Hz, 1H), 3.71 (d, J = 133 Hz, 6H), 2.94 (d, $^{1}J_{CH}$ = 131.9 Hz, 3H). ^{13}C NMR ($CDCl_3$, 100 MHz) 200.5, 160.4, 138.0, 137.9, 134.6, 133.0, 130.1, 126.5, 125.9, 121.9, 121.8, 119.6, 117.0, 112.0, 109.6, 108.1, 33.2, 29.7. LCMS (ESI) calc'd for $[C_{24}{}^{13}C_3H_{22}N_3ONa]^+$($[M+Na]^+$): *m/z* 431.46, found 431.2.

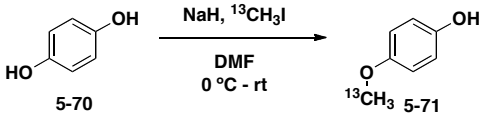
5.7.6 Synthesis of non–indole substrates



To a solution of benzhydrol (586 mg, 3.18) in THF (8 mL) was added sodium hydride (60% suspension in mineral oil, 789 mg, 19.7 mmol). After 30 min, methyl iodide (500 mg, 3.50 mmol) was added drop–wise. After the completion of the reaction (TLC- 15% ethyl acetate in hexane), ice was added to the reaction mixture and then extracted with ether. The organic layer was dried under reduced pressure and purified via flash chromatography on silica gel (5% ethyl acetate in hexane) to give 609 mg clear oil (97% yield).

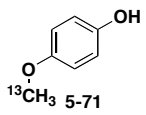


^{1}H NMR ($CDCl_3$, 400 MHz) 7.25 (d, J = 8.5 Hz, 4H), 7.21 (d, J = 7.9 Hz, 4H), 7.15 (dd, J = 5.7, 8.0 Hz, 2H), 5.14 (d, J = 3.4 Hz, 1H), 3.28 (d, J = 142 Hz, 3H). ^{13}C NMR ($CDCl_3$, 100 MHz) 57.28 (^{13}C-label)



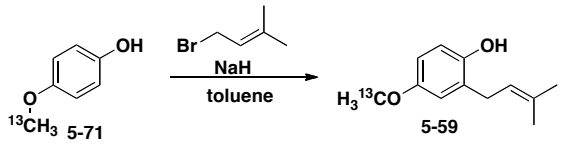
To a solution of hydroquinone (694 mg, 6.3 mmol) in dry DMF (63 mL) at 0 ºC was added K_2CO_3 (827 mg, 6.0 mmol). The reaction mixture slowly turned to a bright orange. Methyl iodide (1.00 g, 7.0 mmol) was added to the reaction mixture drop–wise. After the reaction was complete (by TLC- 30% ethyl acetate in hexane), saturated ammonium chloride was added until

the pH of the reaction mixture reached 7.5. The aqueous mixture was extracted with ethyl acetate and dried under reduced pressure. The crude material was purified via flash chromatography on silica gel (gradient elution 10% ethyl acetate in hexane to 30% ethyl acetate in hexane) to give 41.80 mg (x % yield) of the desired product as yellow oil. The majority of the material was isolated as the di–methylated product.

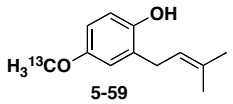


5-71

^{1}H NMR ($CDCl_3$, 600 MHz) 6.78 (d, J = 9.2 Hz, 2H), 6.75 (d, J = 9.0 Hz, 2H), 4.90 (bs, 1H), 3.75 (d, $^1J_{CH}$ = 145 Hz, 3H). ^{13}C NMR ($CDCl_3$, 100 MHz) 55.7 (^{13}C label)

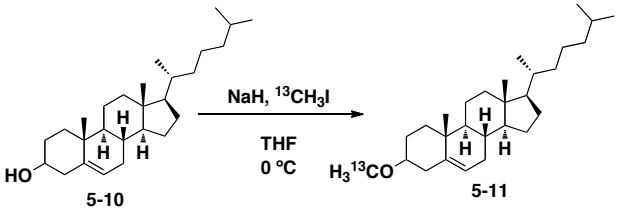


To a solution of 4–methoxyphenol (42 mg, 0.34 mmol) in dry toluene (1.0 mL) at 0 ºC was added sodium hydride (14.8 mg, 0.37 mmol). The reaction mixture became bright blue. After 90 min, 1–bromo–2–methyl–2–butene (55 mg, 0.37 mmol) was added drop–wise. After completion of the reaction (by TLC 15% ethyl acetate in hexane) saturated ammonium chloride was added to the reaction mixture. The aqueous material was extracted with ether and dried under reduced pressure. The crude material was purified via flash chromatography on silica gel (10% ethyl acetate in hexane) to give 25.4 mg of the desired product as clear oil (39% yield).



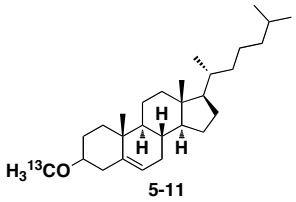
5-59

^{1}H NMR ($CDCl_3$, 600 MHz) 6.71 (d, J = 8.6 Hz, 1H), 6.67 (d, J = 2.7 Hz, 1H), 6.63 (dd, J = 2.8, 8.5 Hz, 1H), 5.29 (dd, J = 7.0, 7.0 Hz, 1H), 4.80 (bs, 1H), 3.73 (d, $^{1}J_{CH}$ = 143 Hz, 3H), 3.30 (d, J = 7.1 Hz, 2H), 1.75 (s, 6H). ^{13}C NMR ($CDCl_3$, 100 MHz)



5-10 5-11

To a solution of cholesterol (500 mg, 1.29 mmol) in THF (5 mL) at 0 ºC was added sodium hydride (155 mg, 3.87 mmol). After one hour, methyl iodide (203 mg, 1.42 mmol) was added drop–wise. After completion of the reaction (by TLC 30% ethyl acetate in hexane) saturated ammonium chloride was added. The aqueous mixture was extracted with ether and concentrated under reduced pressure to give 2.38 g white solid (94% yield).



5-11

^{1}H NMR ($CDCl_3$, 400 MHz) 5.34 (t, J = 1.68 Hz, 1H), 3.34 (d, $^{1}J_{CH}$ = 141 Hz, 3H), 3.04 (m, 1H), 2.37 (d, J = 12 Hz, 1H), 2.22 (dd, J = 6.2, 9.2 Hz, 1H), 2.13 (t, J = 13.6 Hz, 1H), 1.98 – 1.83 (m, 6H), 1.57 (s, 3H), 1.51 – 1.45 (m, 6H), 1.32–1.31 (m, 3H), 1.24 (m, 4H), 1.14 – 1.03

(m, 8H), 1.02–0.96 (m, 3H), 0.98 (s, 3H), 0.89 (d, J = 6.4 Hz, 3H), 0.85 (d, J = 6.6 Hz, 6H), 0.66 (s, 3H). ^{13}C NMR ($CDCl_3$, 100 MHz)

5.8 References

1. Silverstein, R. M., Webster, F. X., Kiemle, D. J., *Spectrometric identification of organic compounds*. John Wiley and Sons, 2005. 7. Print.

2. Rogers, J. L., MacMillan, J. B., *J. Am. Chem. Soc.* **2012**, *134*. 12378–12381. A labeled substrate approach to discovery of biocatalytic reactions: a proof of concept transformation with N–methylindole.

3. Hirose, T., Sunazuka, T., Yamamoto, D., Kojima, N., Shirahata, T., Harigaya, Y., Kuwajima, I., Omura, S., *Tetrahedron*. **2005**, *61*, 6015–6039. Determination of the absolute stereochemistry and asymmetric total synthesis of madindolines A and B: a practical improvement to a second–generation approach from the first–generation.

4. Alvarez, R. G., Hunter, I. S., Suckling, C. L., Thomas, M., Vitinius, U., *Tetrahedron,* **2001**, *57*, 8581–8587. A novel biotransformation of benzofurans and related compounds catalysed by a chloroperoxidase.

5. Kumar, C. N. S. S. P., Devi, C. L., Rao, V. J., Palaniappan, S., *Synlett*, **2008**, *13*, 2023–2027. Use of pyridinium chlorochromate and reusable polyaniline salt catalyst combination for the oxidation of indoles.

6. Miller, K. A., Tsukamoto, S., Williams, R. M., *Nature Chemistry*, **2009**, *1*, 63–68. Asymmetric total synthesis of (+)– and (–)–versicolamide B and biosynthetic implications.

7. Pettersson, M., Knueppel, D., Martin, S. F., *Org. Lett*. **2007**, *9*, 4623–4626. Concise, stereoselective approach to the spirooxindole ring system of cintinadin A.

8. Bai, C., Jiang, Y., Hu, M., Li, S., Zhai, Q., *Catal. Lett.* **2009**, *129*, 457–461. Improvement of chloroperoxidase activities by chitosan and thioglycolic acid.

9. Medici, R., Garaycoechoa, J. I., Detorre, L. A., Iribarran, A. M., Lewkowicz, E. S., *Biotechnol. Lett.* **2011**, *33*, 1999–2003. Biocatalyzed halogenation of nucleobase analogs.

10. Syu, S., Lee, Y–T., Jang, Y–J., Lin, W., *Org. Lett*, **2011**, *13*, 2970–2973. Preparation of functional benzofurans, benzothiophenes and indoles using ester, thioester, and amide via intramolecular Wittig reactions.

11. Okuro, K., Gurnham, J., Alper, H., *J. Org. Chem.*, **2011**, *76*, 4715–4720. Ionic diamine rhodium complex catalyzed reductive N–heterocylization of 2–nitrovinylarenes.

12. Solomons, T. W.G., Fryhle, C. B., *Organic Chemistry.* 10e. Hoboken, NJ: Wiley and Sons, 2011. 4. Print.

13. Petrier, C., Lavaitte, S., Morat, C., *J. Org. Chem.* **1990**, *55*, 1664–1667. Properties of the zinc–nickel chloride–deuterium oxide system: a simple method for deuterium addition to carbon-carbon and carbon–oxygen double bonds.

14. Moriarty, R. M.; David, N. E., Mahmood, N. A., WO2012021133 A1 2012. Kythera Biopharmaceuticals, Inc. Synthetic bile acid compositions.

15. Labrie, F., Singh, S., Gauthier, S., Frechette, Y., Chenard, S., Breton, R., WO 2005066194. English. Preparation of helix 12 directed steroids for the treatment of androgen related diseases.

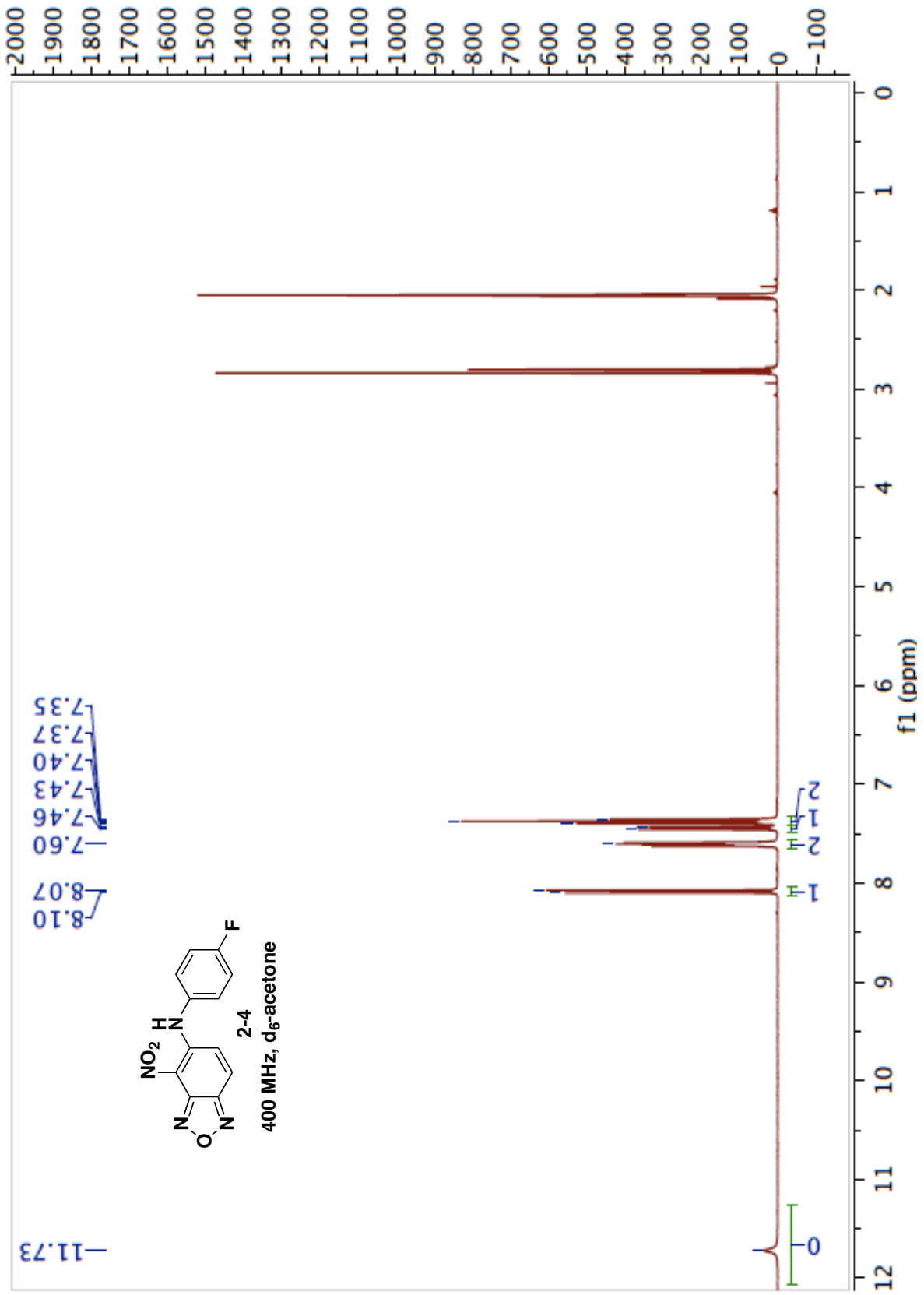
16. Hillisch, A., Boidol, W., Schwede, W., Esperling, P., Sauer, G., Hegele–Hartung, C., Kollenkirchen, U., Fritzemeier, K–H., WO 2001032680 A2 20010510. German. Preparation of 18 nor–steroids as selectively active estrogens.

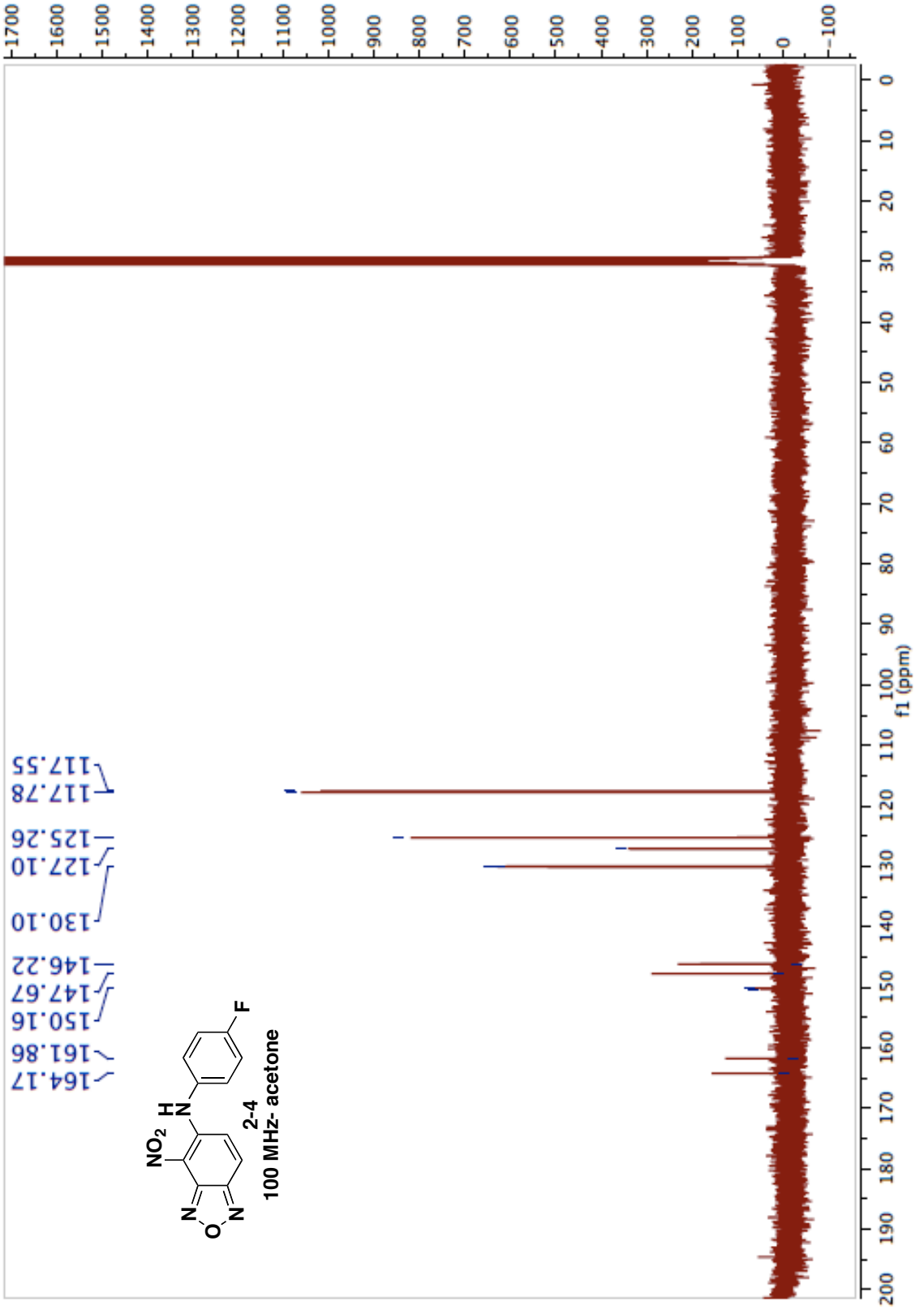
17. Rhee, S–W., Malerich, J. P., Tanga, M. J., Lee, J–J. *J. Labelled Comp. Radiopharm.* **2012**, *55*, 186–187. Synthesis of 14–C acanthoic acid.

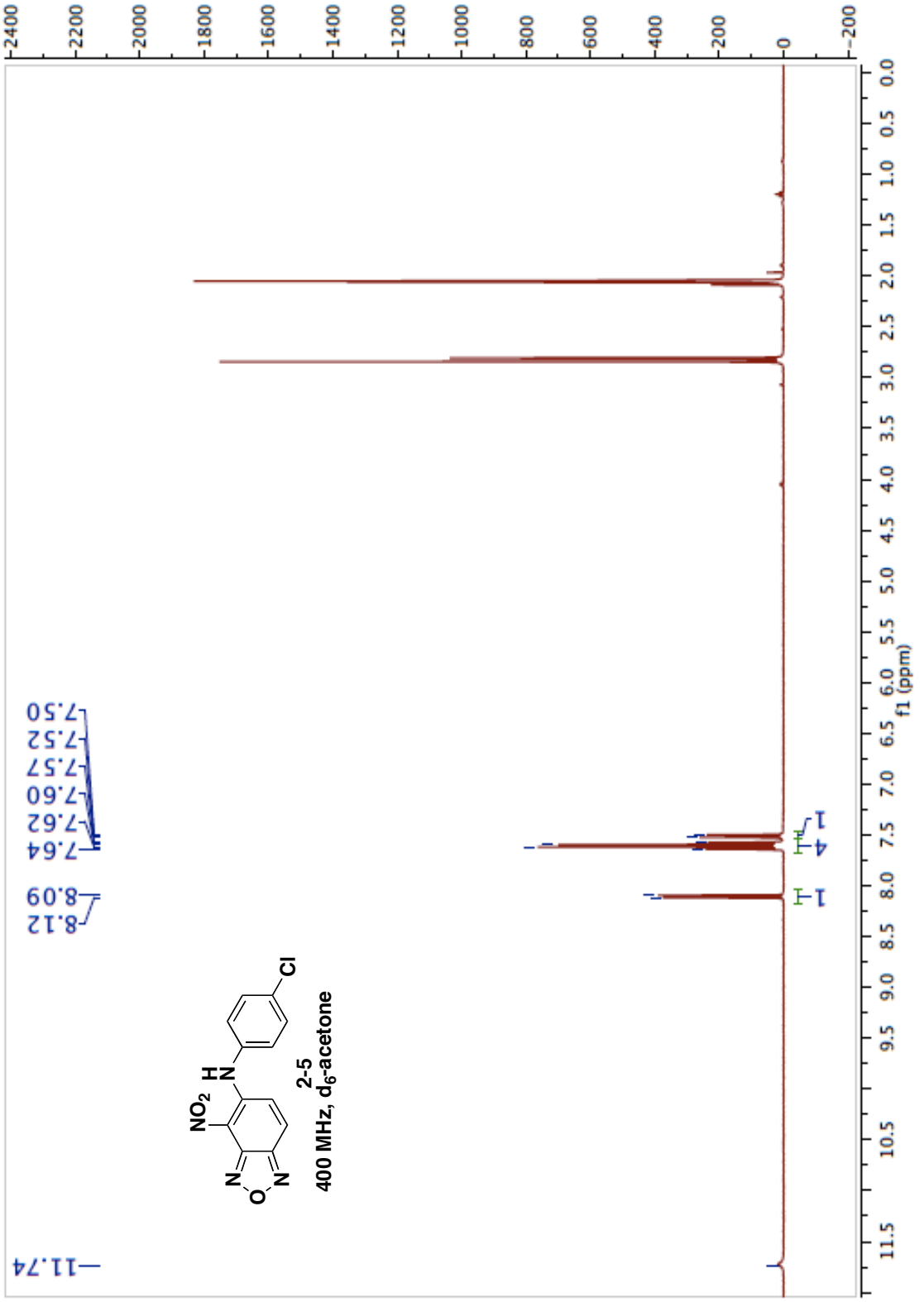
18. Temple, J. S., Riediker, M., Schwartz, J., *J. Am. Chem. Soc.* **1982**, *104*, 1310–1315. Regiocontrolled coupling of (n–allyllic)palladium complexes with organozirconium species.

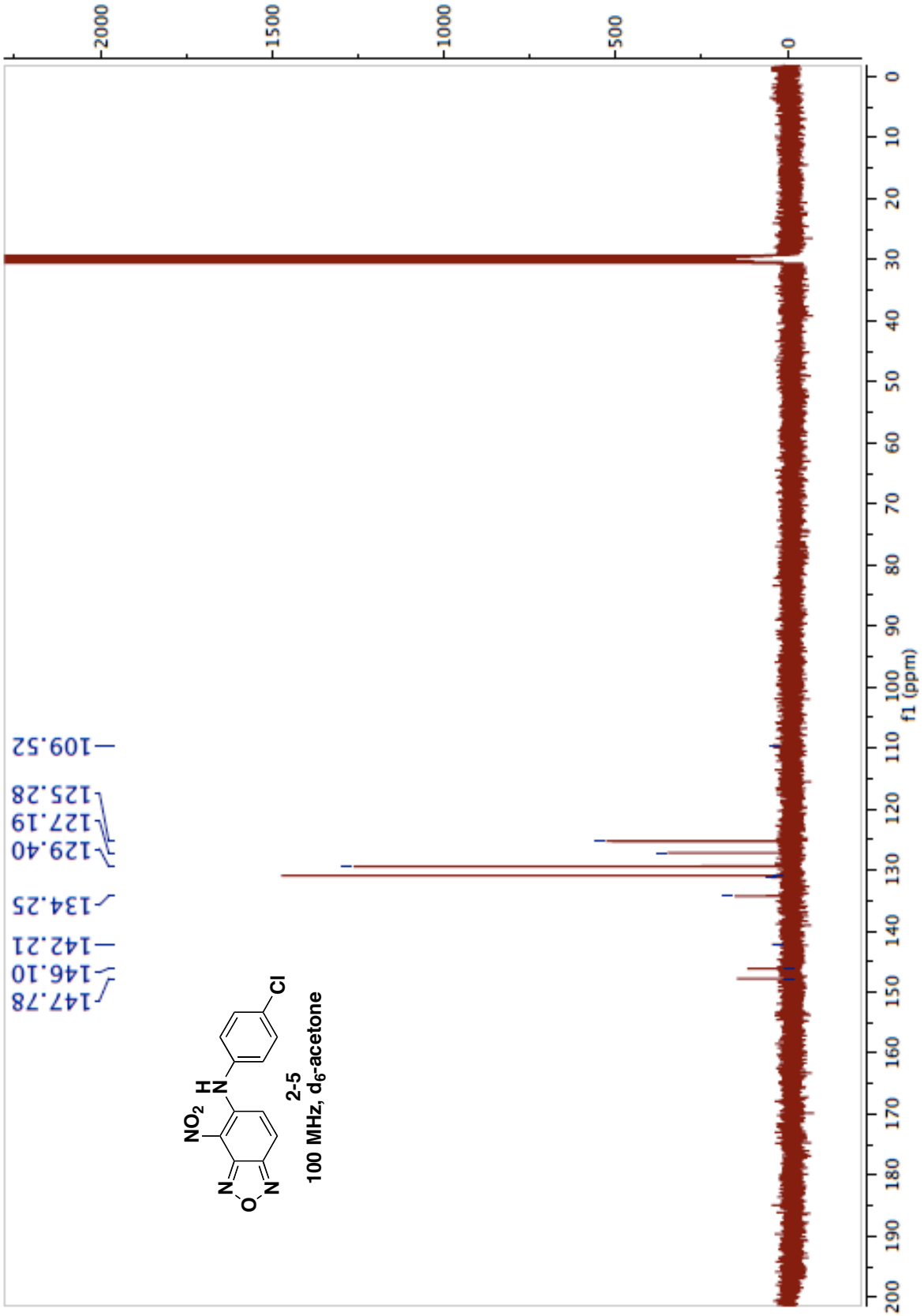
19. ^{14}C can be detected in tissues at a concentration of 10^{-18} moles per tissue. Dormer, P. *The Histochemical Journal*, **1981**, *13*, 161–171. Quantitative carbon–14 autoradiography at the cellular level: Principles and application for cell kinetic studies.

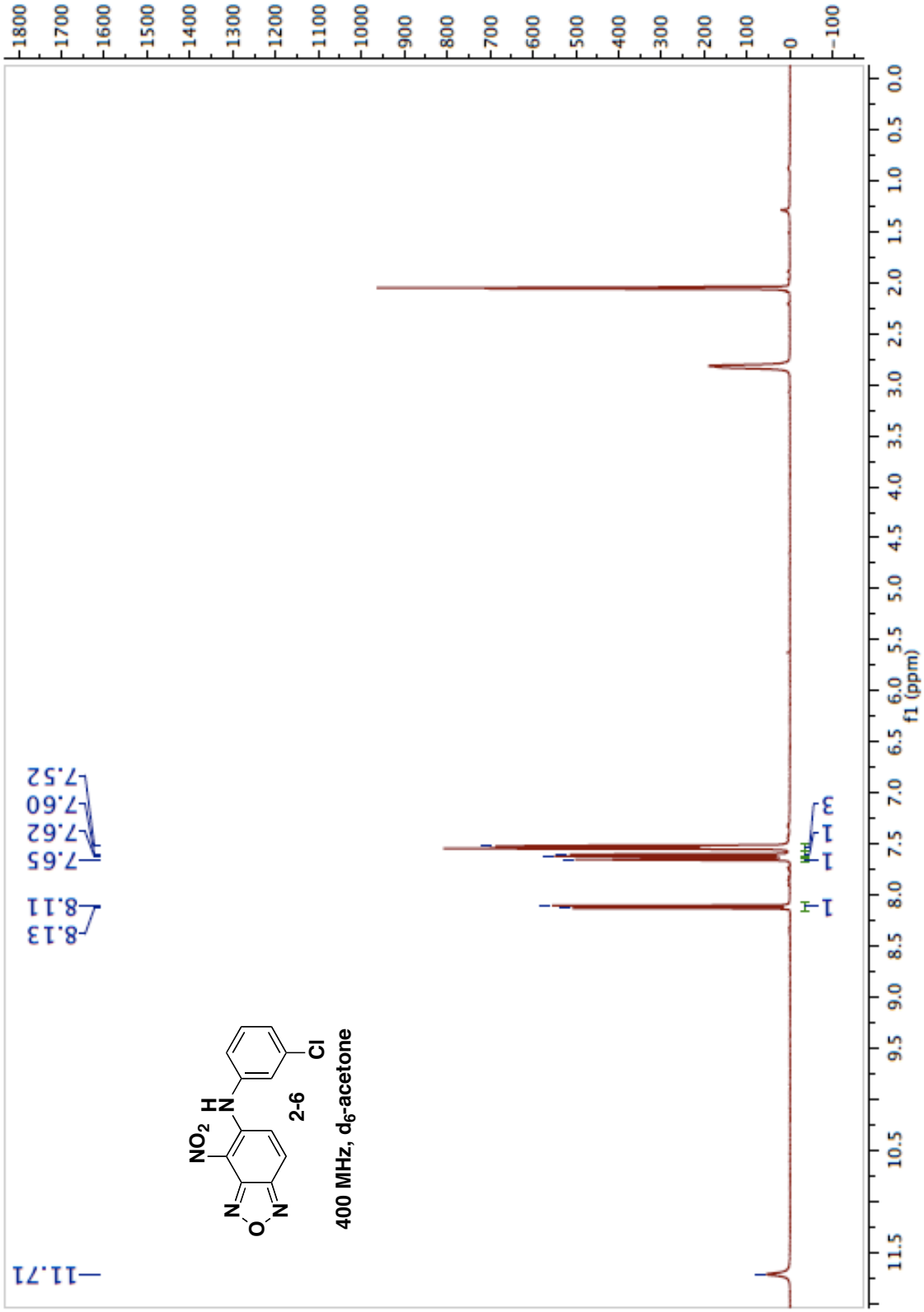
APPENDICES

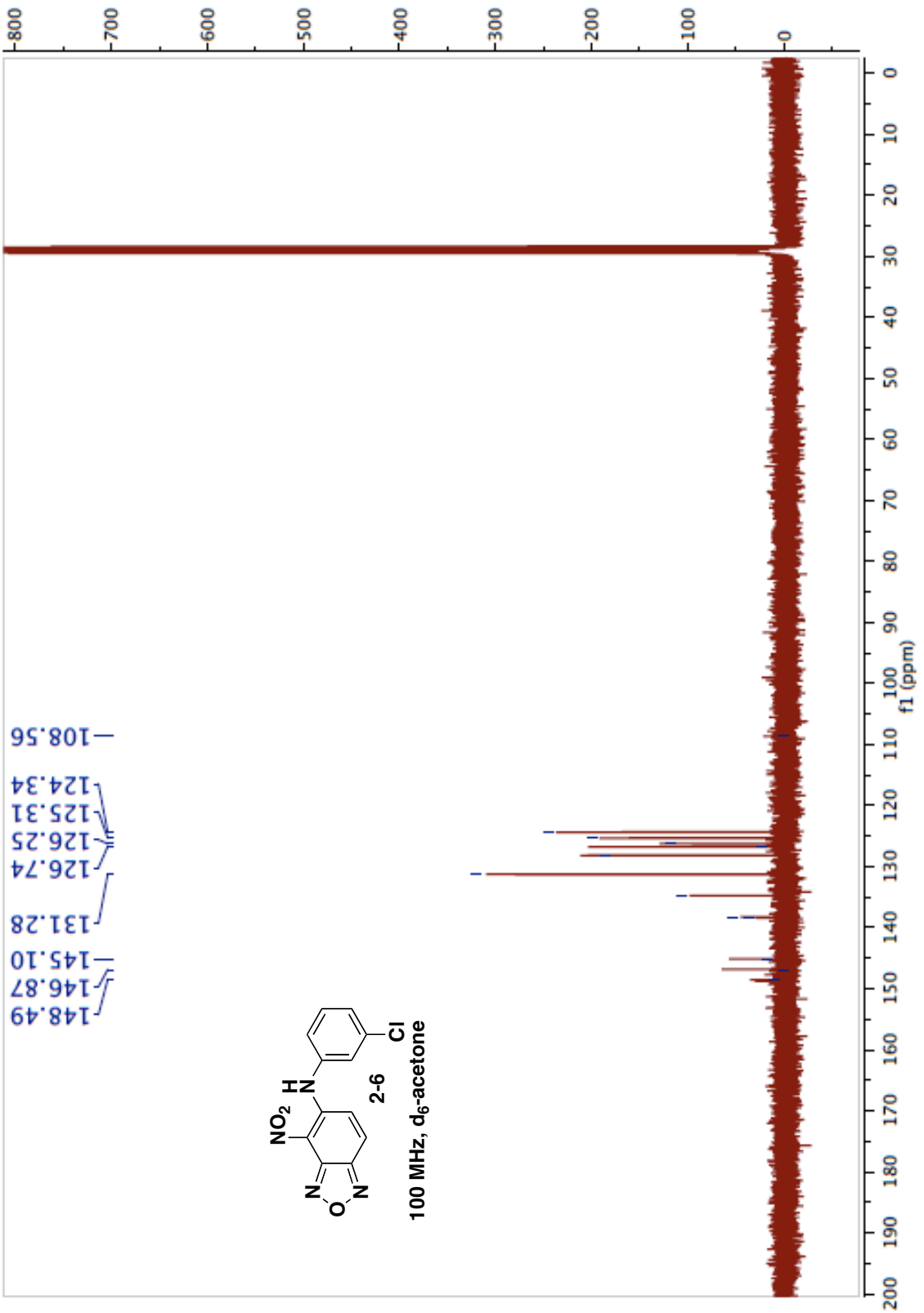

11.73
8.10
8.07
7.60
7.46
7.43
7.40
7.37
7.35
2-4
400 MHz, d6-acetone
f1 (ppm)

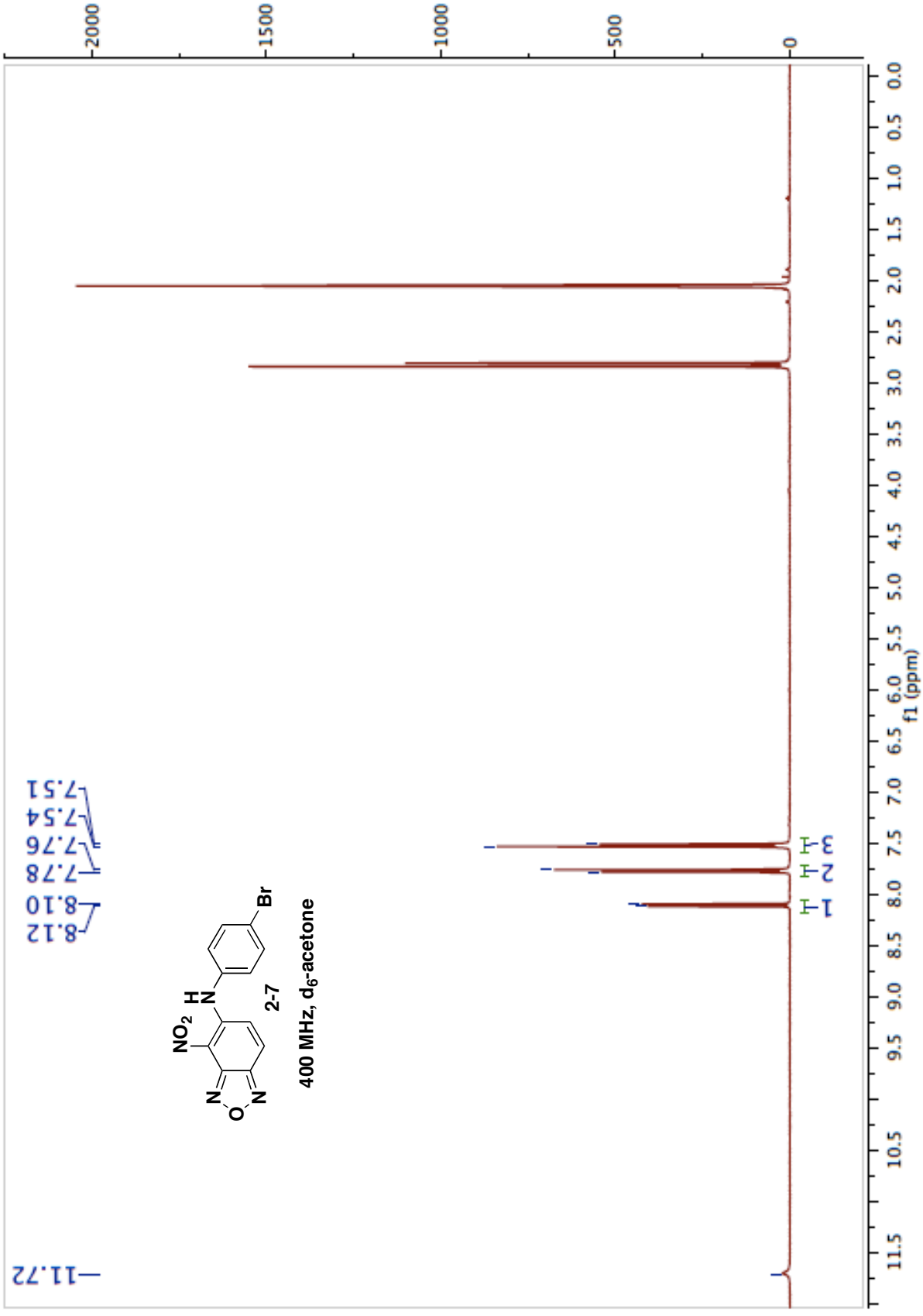

2-4
100 MHz- acetone
NO2
H
N
F
N
O
N
164.17
161.86
150.16
147.67
146.22
130.10
127.10
125.26
117.78
117.55
f1 (ppm)

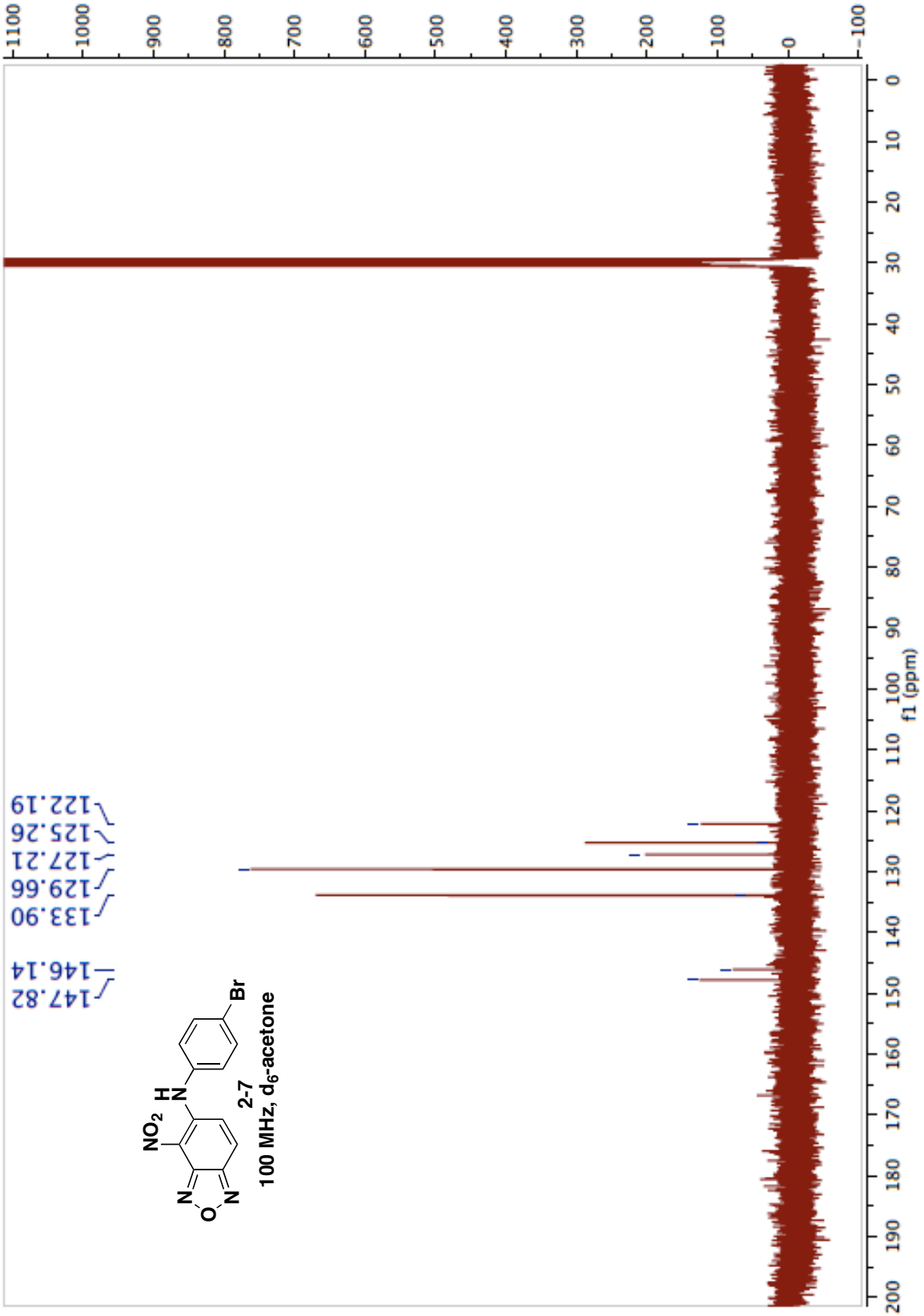

11.74
8.12
8.09
7.64
7.62
7.60
7.57
7.52
7.50
1
4
1
2-5
400 MHz, d6-acetone
f1 (ppm)

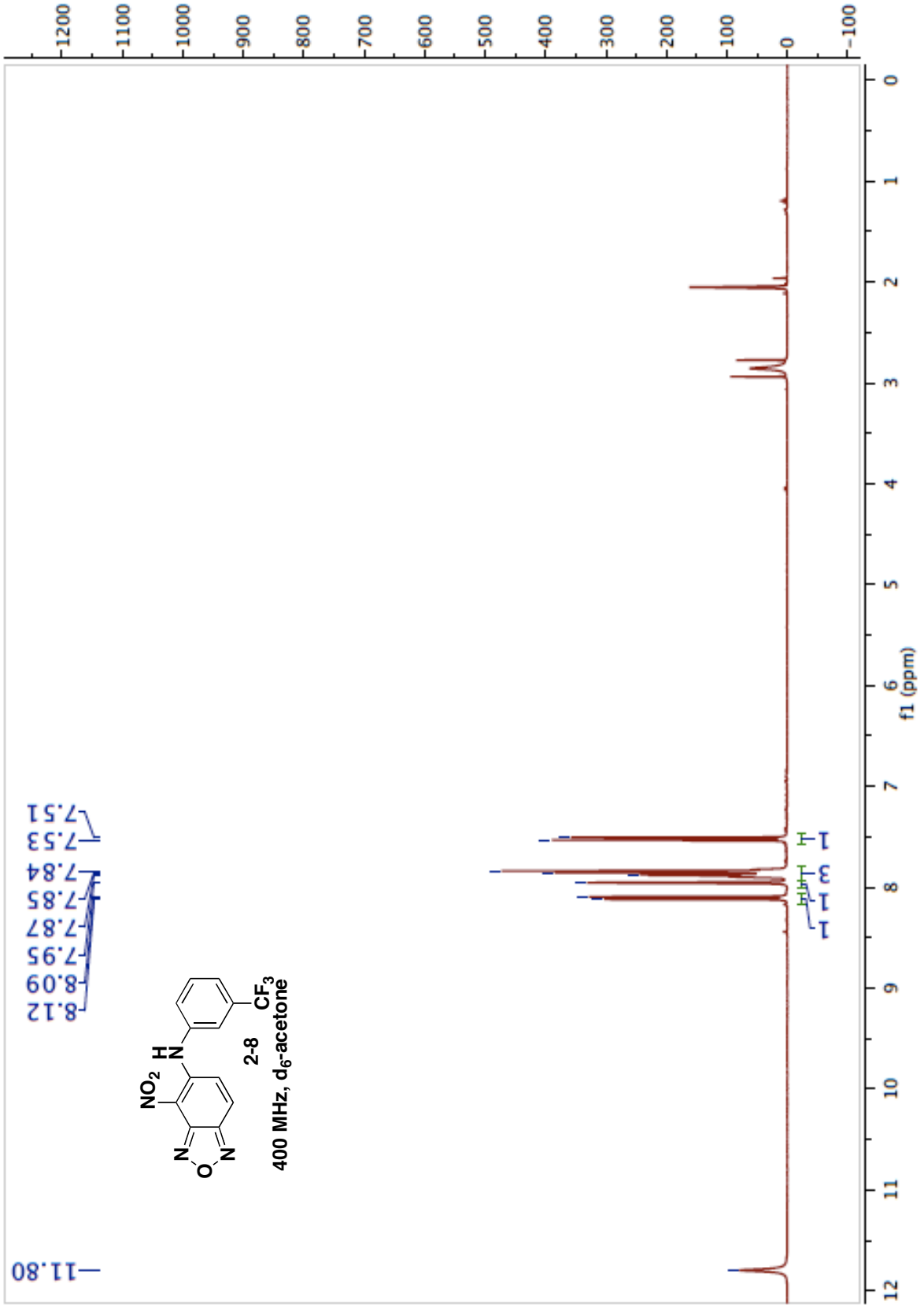

2000
1500
1000
500
0
f1 (ppm)
109.52
125.28
127.19
129.40
134.25
142.21
146.10
147.78
2-5
100 MHz, d6-acetone
NO2
H
N
Cl
N
O
N

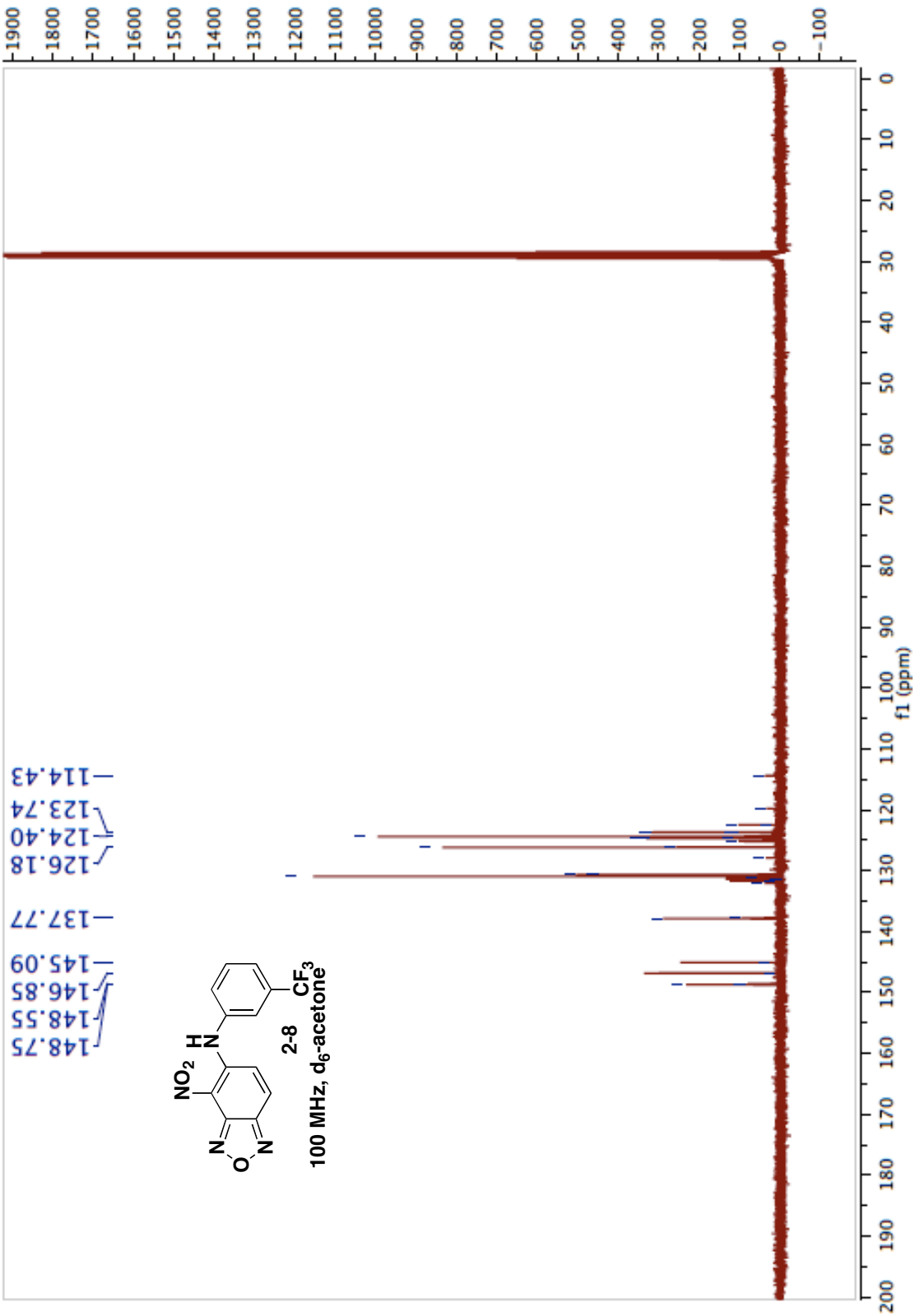

11.71
8.13
8.11
7.65
7.62
7.60
7.52
1
1
3
f1 (ppm)
2-6
400 MHz, d6-acetone

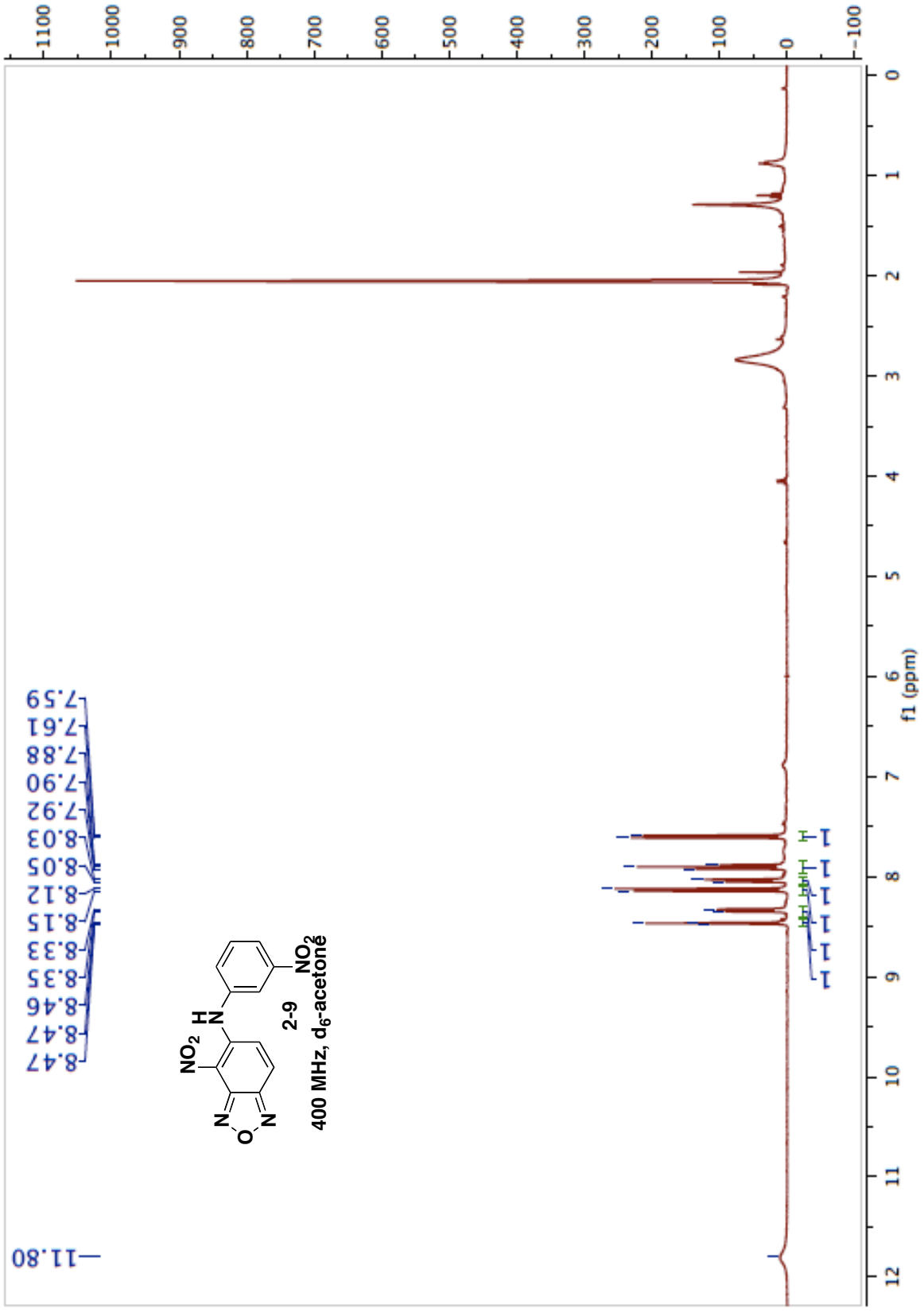

148.49
146.87
145.10
131.28
126.74
126.25
125.31
124.34
108.56
100 MHz, d6-acetone
2-6
NO2
H
N
Cl
f1 (ppm)

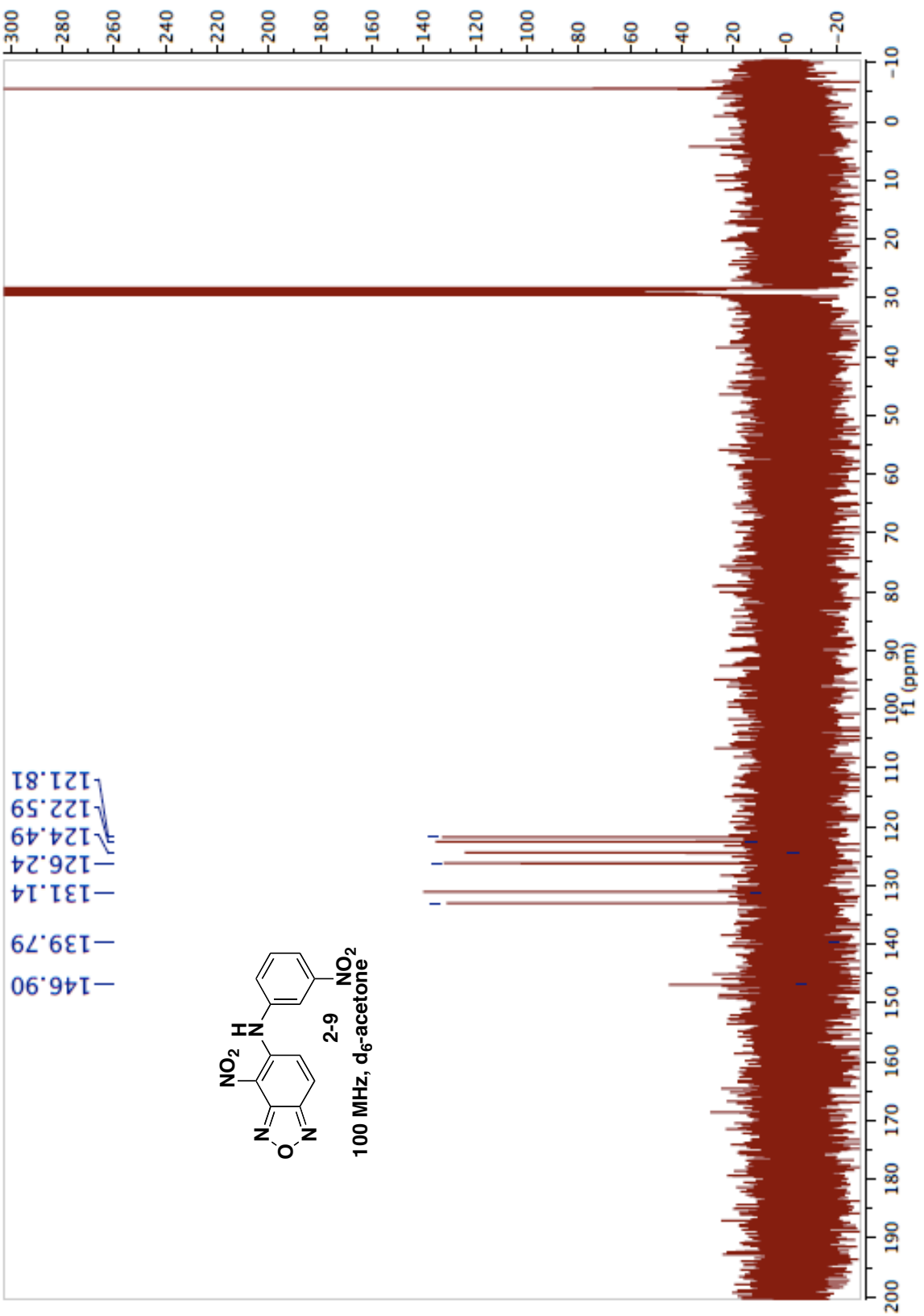

11.72
8.12
8.10
7.78
7.76
7.54
7.51
1
2
3
2-7
400 MHz, d6-acetone
f1 (ppm)

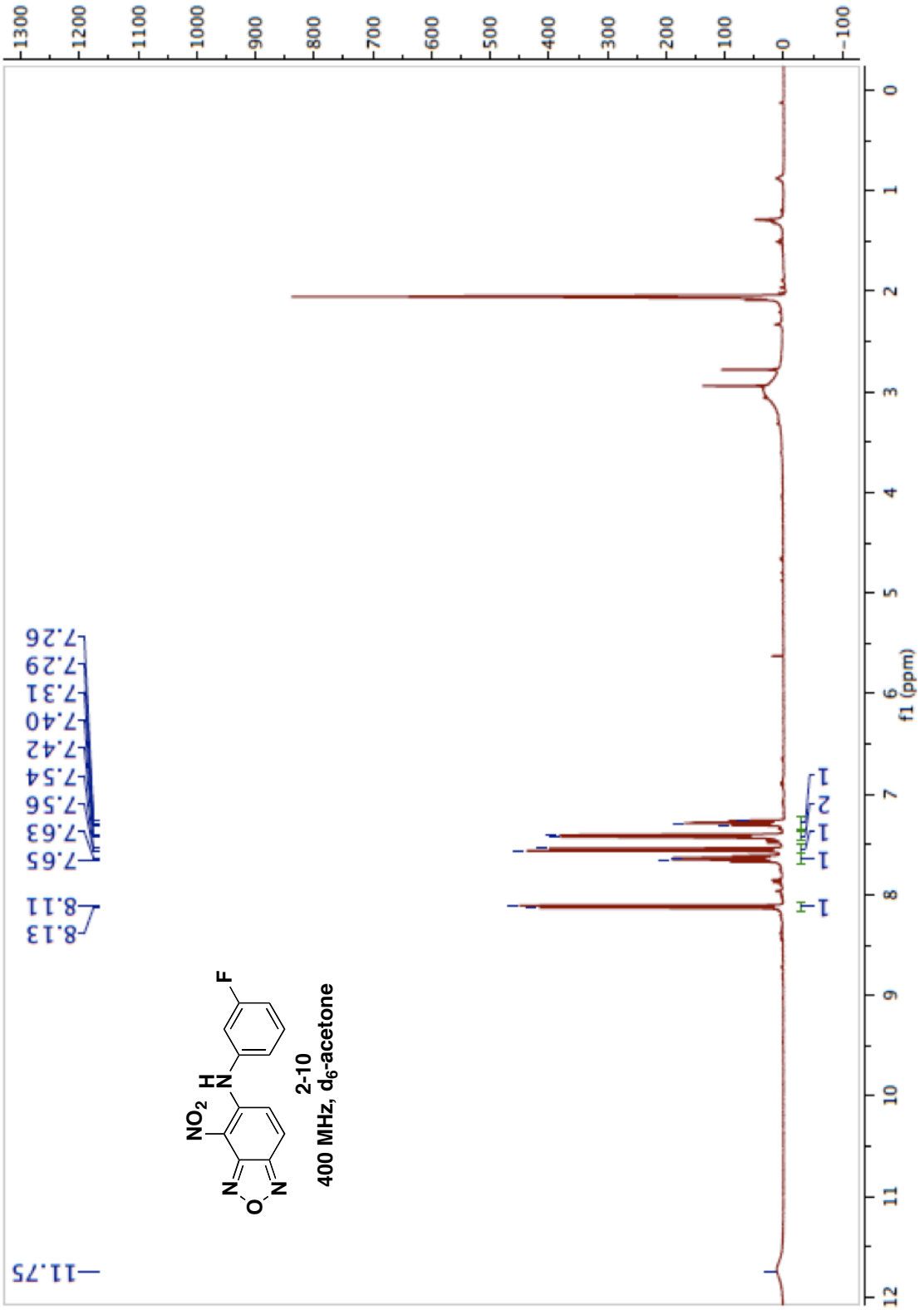

147.82
146.14
133.90
129.66
127.21
125.26
122.19
2-7
100 MHz, d6-acetone
f1 (ppm)

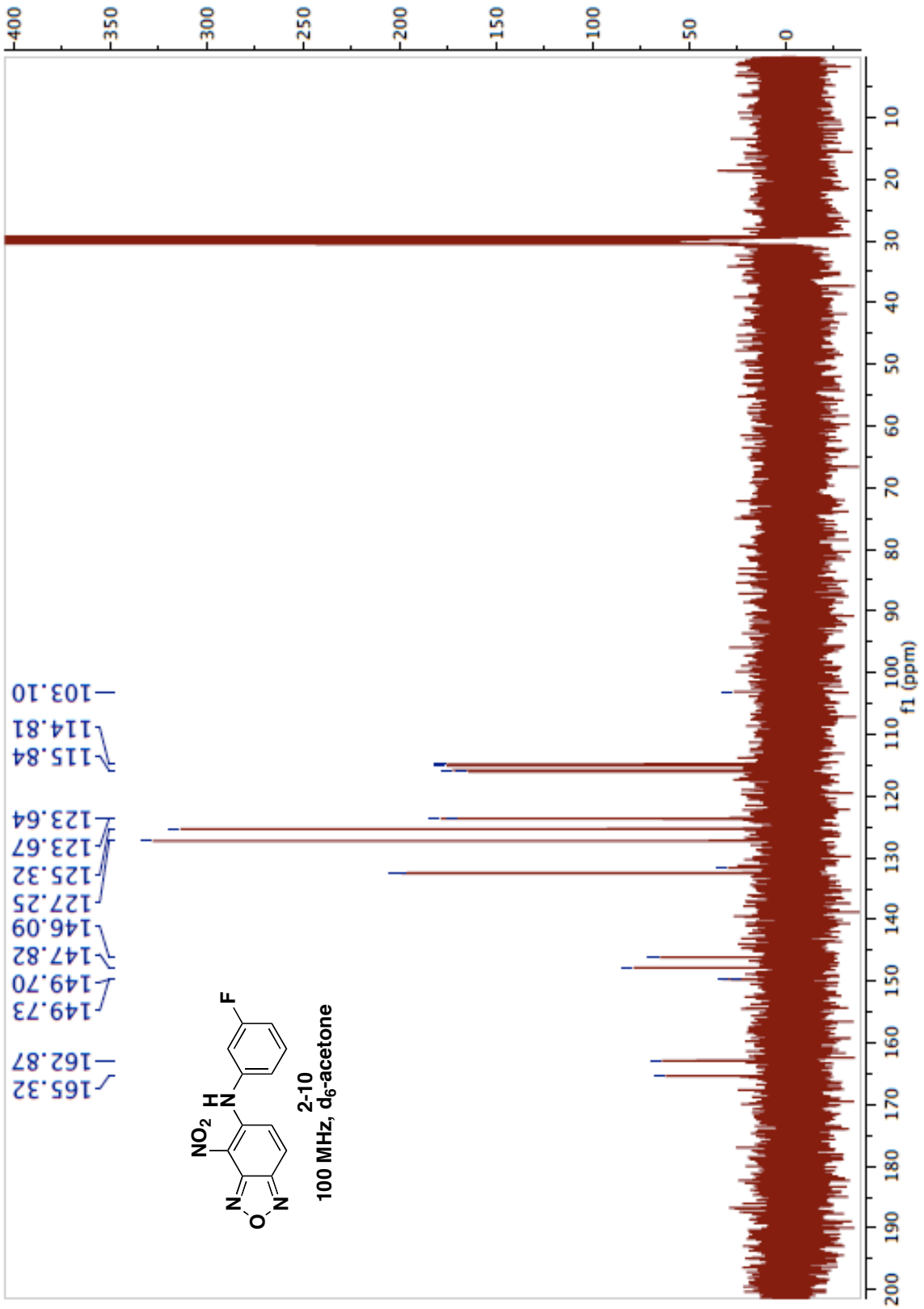

11.80
8.12
8.09
7.95
7.87
7.85
7.84
7.53
7.51
1
3
1
1
f1 (ppm)
2-8
400 MHz, d6-acetone

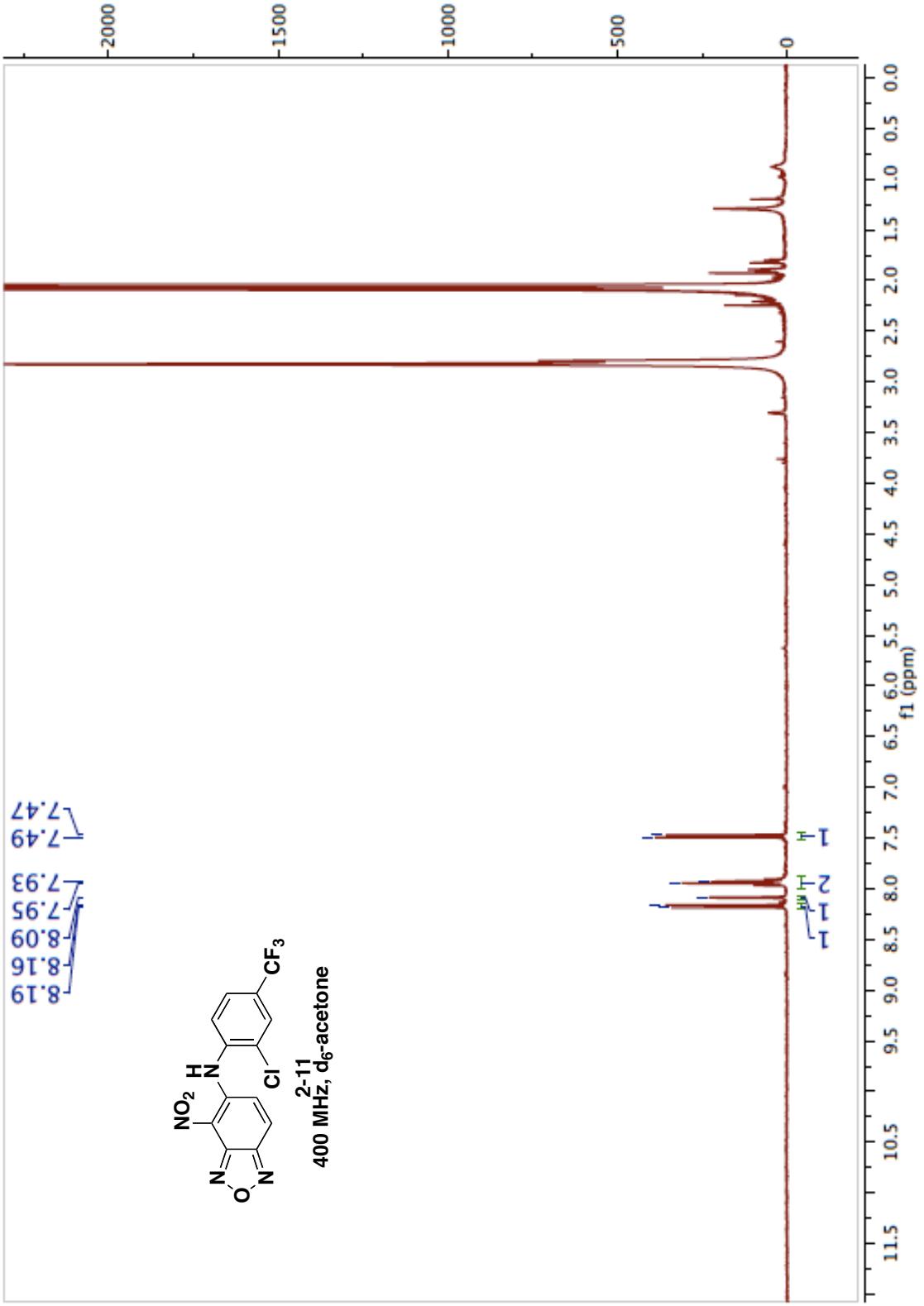

148.75
148.55
146.85
145.09
137.77
126.18
124.40
123.74
114.43
2-8
100 MHz, d6-acetone
f1 (ppm)

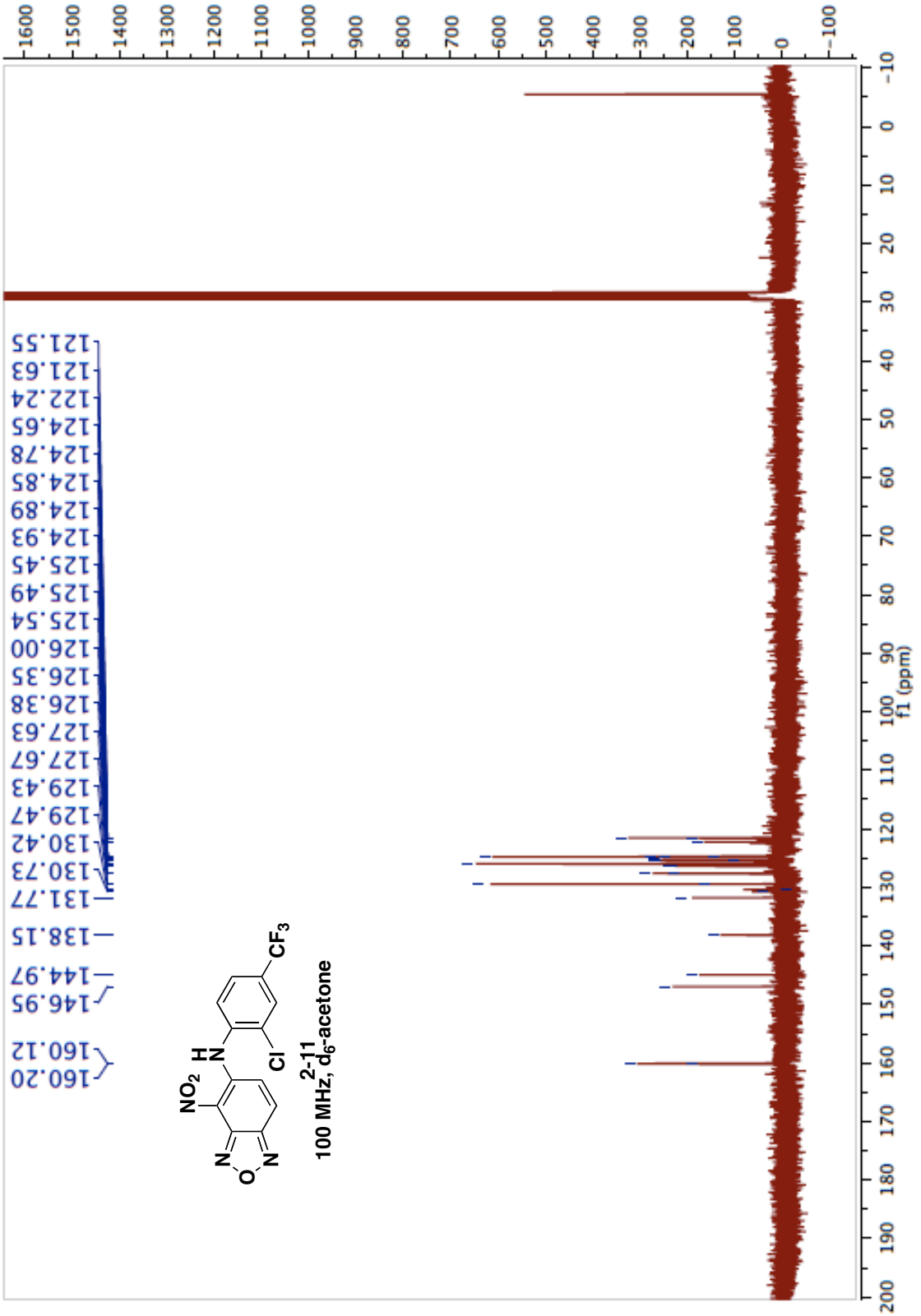

11.80
8.47
8.47
8.46
8.35
8.33
8.15
8.12
8.05
8.03
7.92
7.90
7.88
7.61
7.59
2-9
400 MHz, d6-acetone
f1 (ppm)

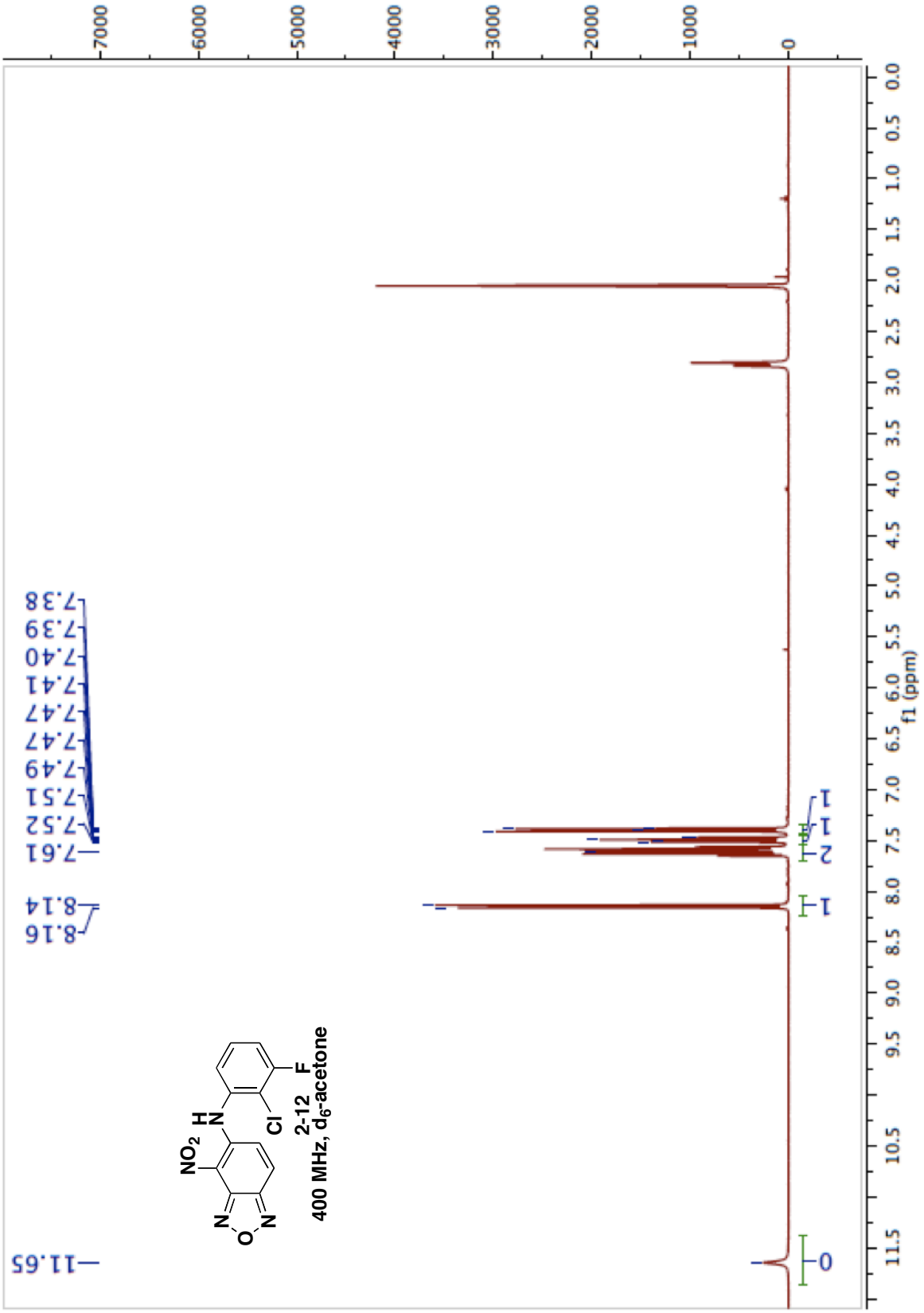

146.90
139.79
131.14
126.24
124.49
122.59
121.81
2-9
100 MHz, d6-acetone
f1 (ppm)

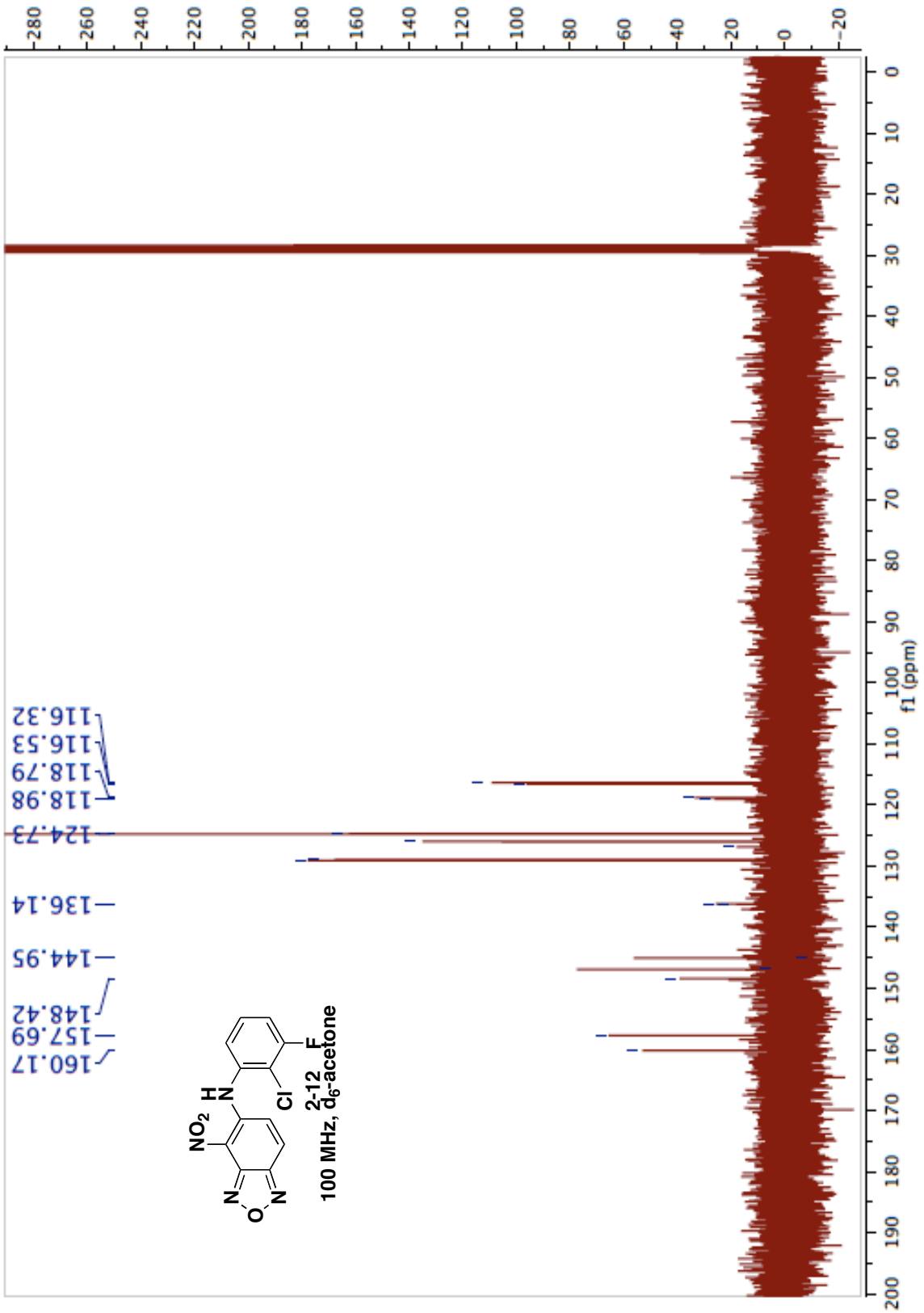

11.75
8.13
8.11
7.65
7.63
7.56
7.54
7.42
7.40
7.31
7.29
7.26
2-10
400 MHz, d6-acetone
f1 (ppm)

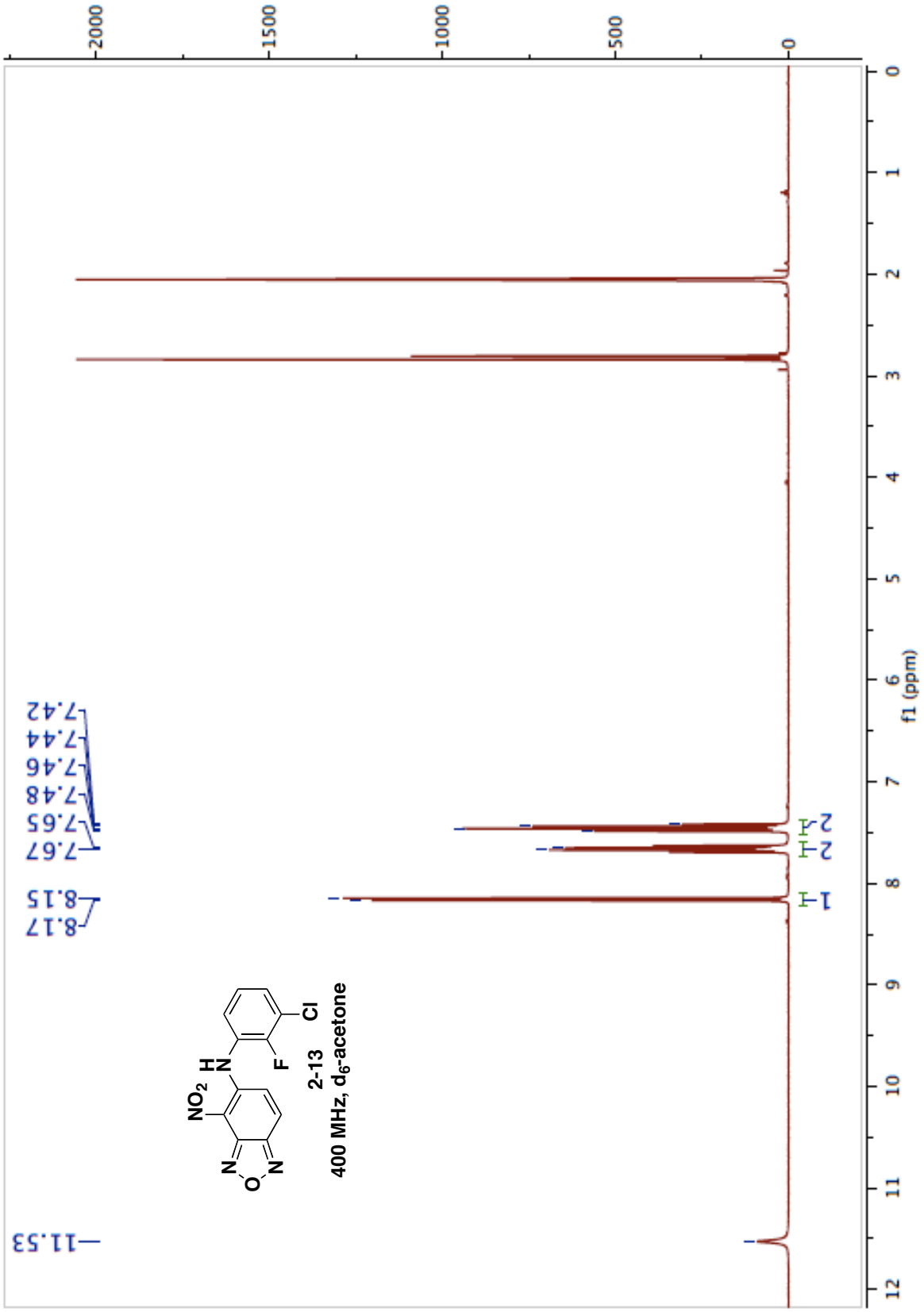

165.32
162.87
149.73
149.70
147.82
146.09
127.25
125.32
123.67
123.64
115.84
114.81
103.10
2-10
100 MHz, d6-acetone
f1 (ppm)

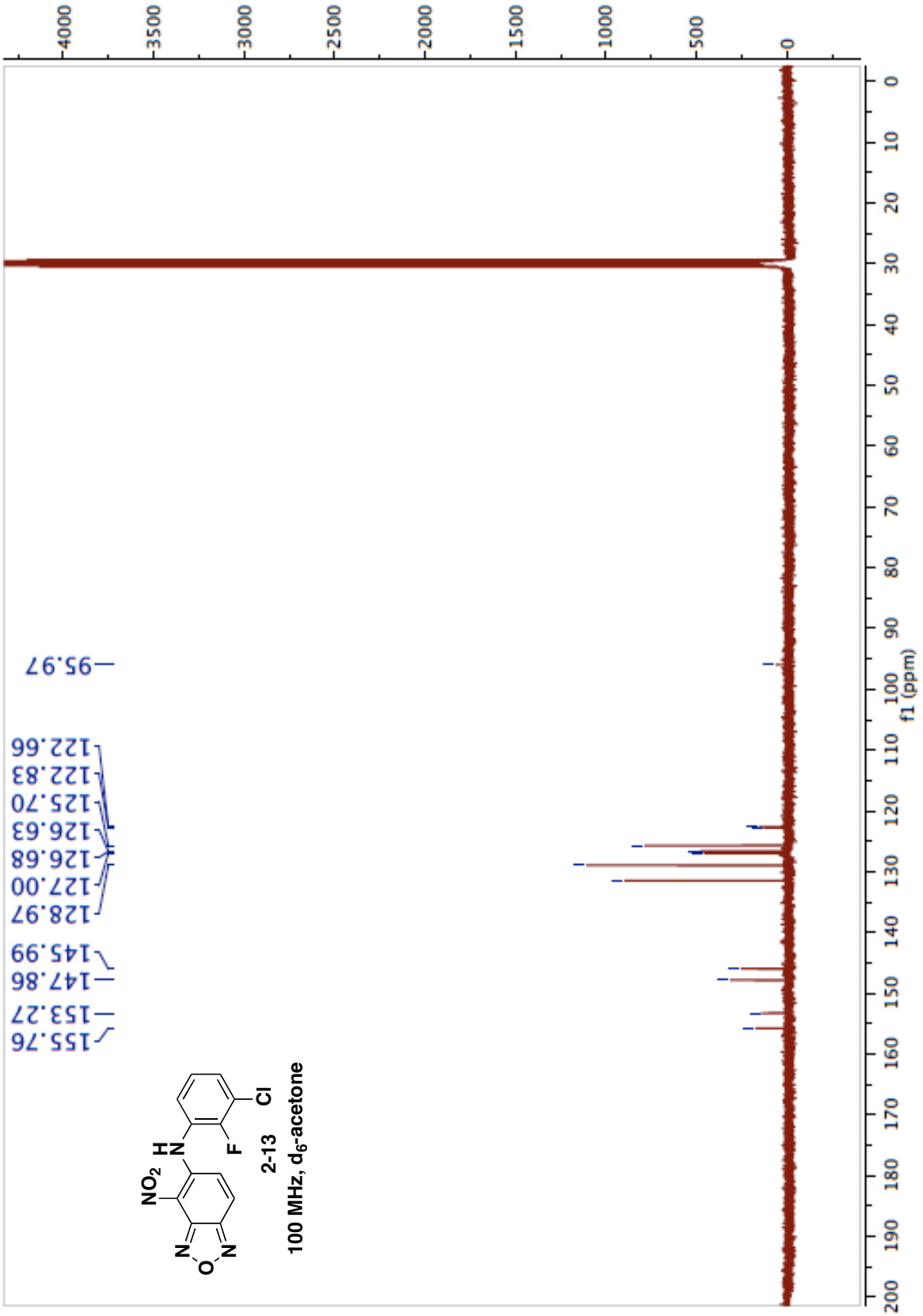

2000
1500
1000
500
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.5
11.5
f1 (ppm)
7.47
7.49
7.93
7.95
8.09
8.16
8.19
1
2
1
1
NO2
H
N
Cl
CF3
N
O
N
2-11
400 MHz, d6-acetone

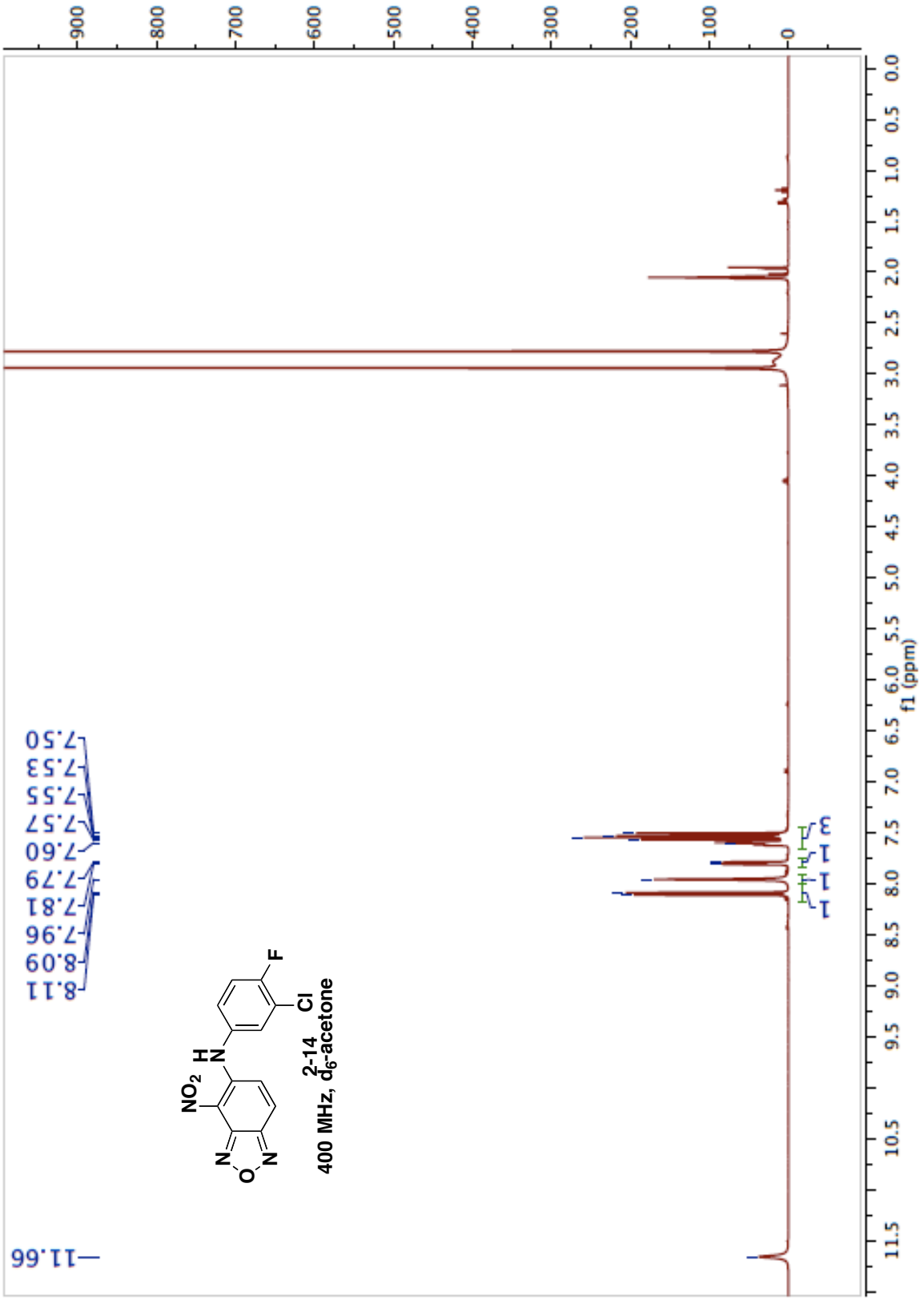

160.20
160.12
146.95
144.97
138.15
131.77
130.73
130.42
129.47
129.43
127.67
127.63
126.38
126.35
126.00
125.54
125.49
125.45
124.93
124.89
124.85
124.78
124.65
122.24
121.63
121.55
NO2
H
N
Cl
CF3
N
O
N
2-11
100 MHz, d6-acetone
f1 (ppm)

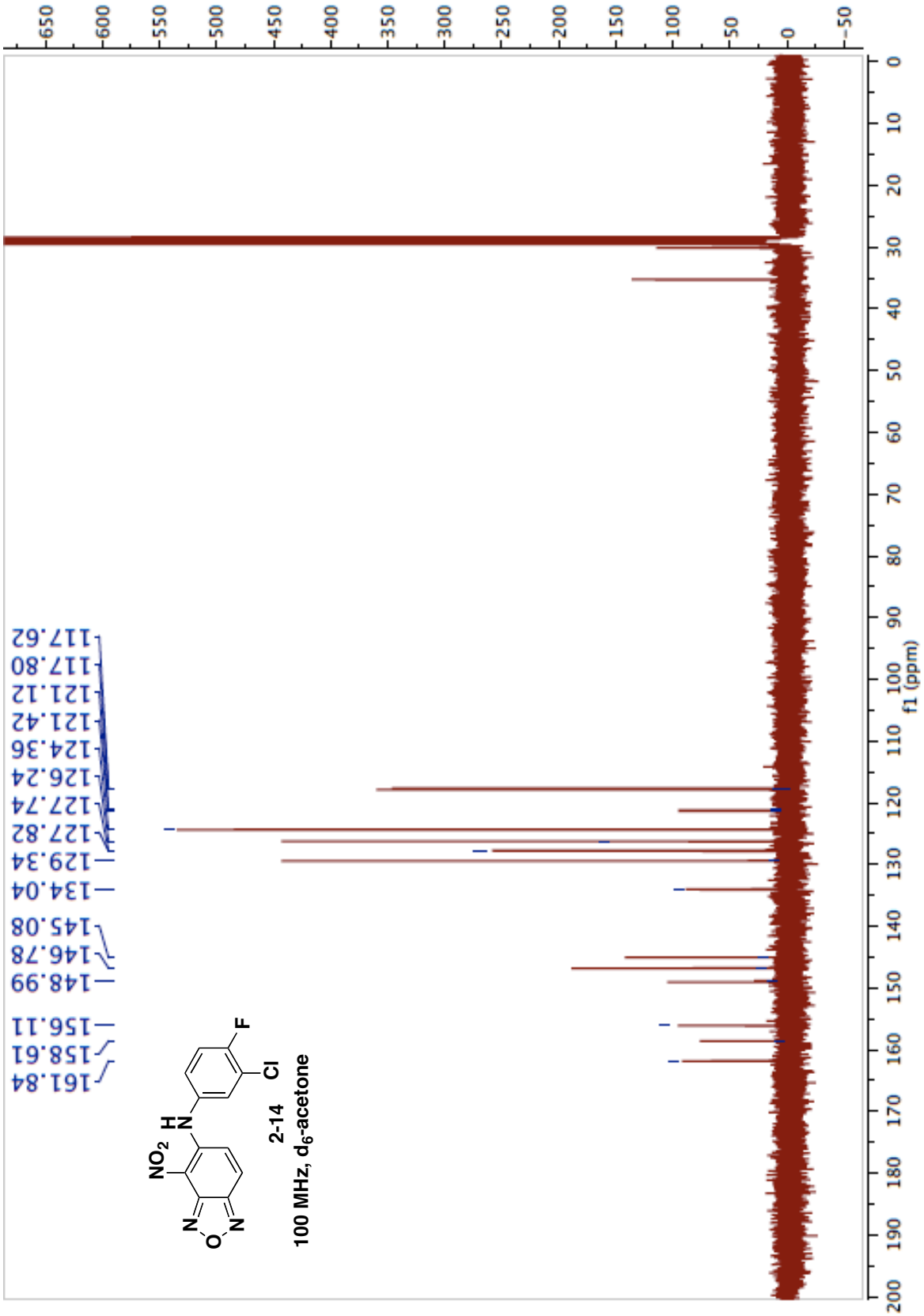

11.65
8.16
8.14
7.61
7.52
7.51
7.49
7.47
7.47
7.41
7.40
7.39
7.38
2-12
400 MHz, d6-acetone
f1 (ppm)

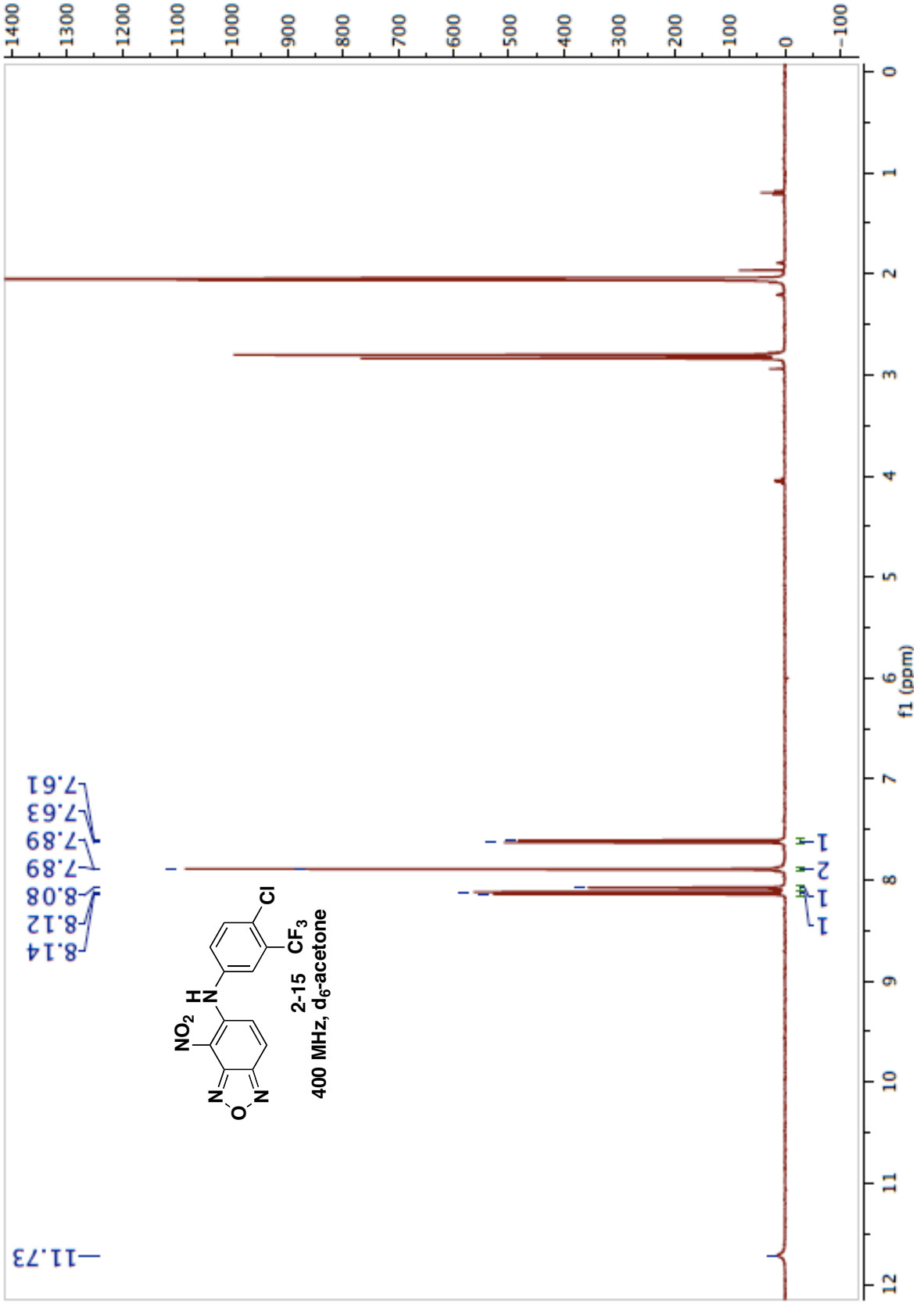

160.17
157.69
148.42
144.95
136.14
124.73
118.98
118.79
116.53
116.32
f1 (ppm)
2-12_F
100 MHz, d6-acetone

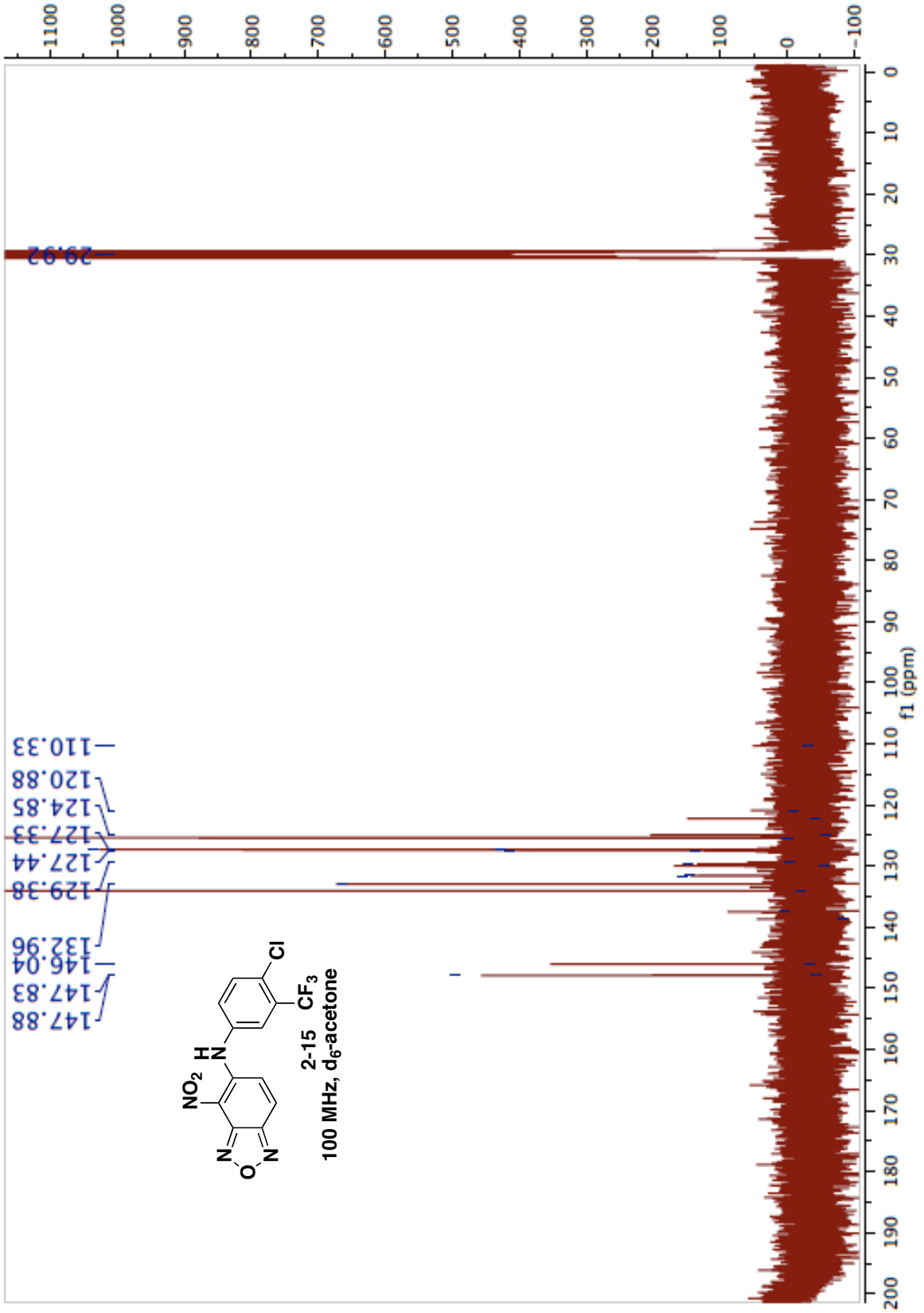

11.53
8.17
8.15
7.67
7.65
7.48
7.46
7.44
7.42
1.00
2.01
2.00
2-13
400 MHz, d6-acetone
f1 (ppm)

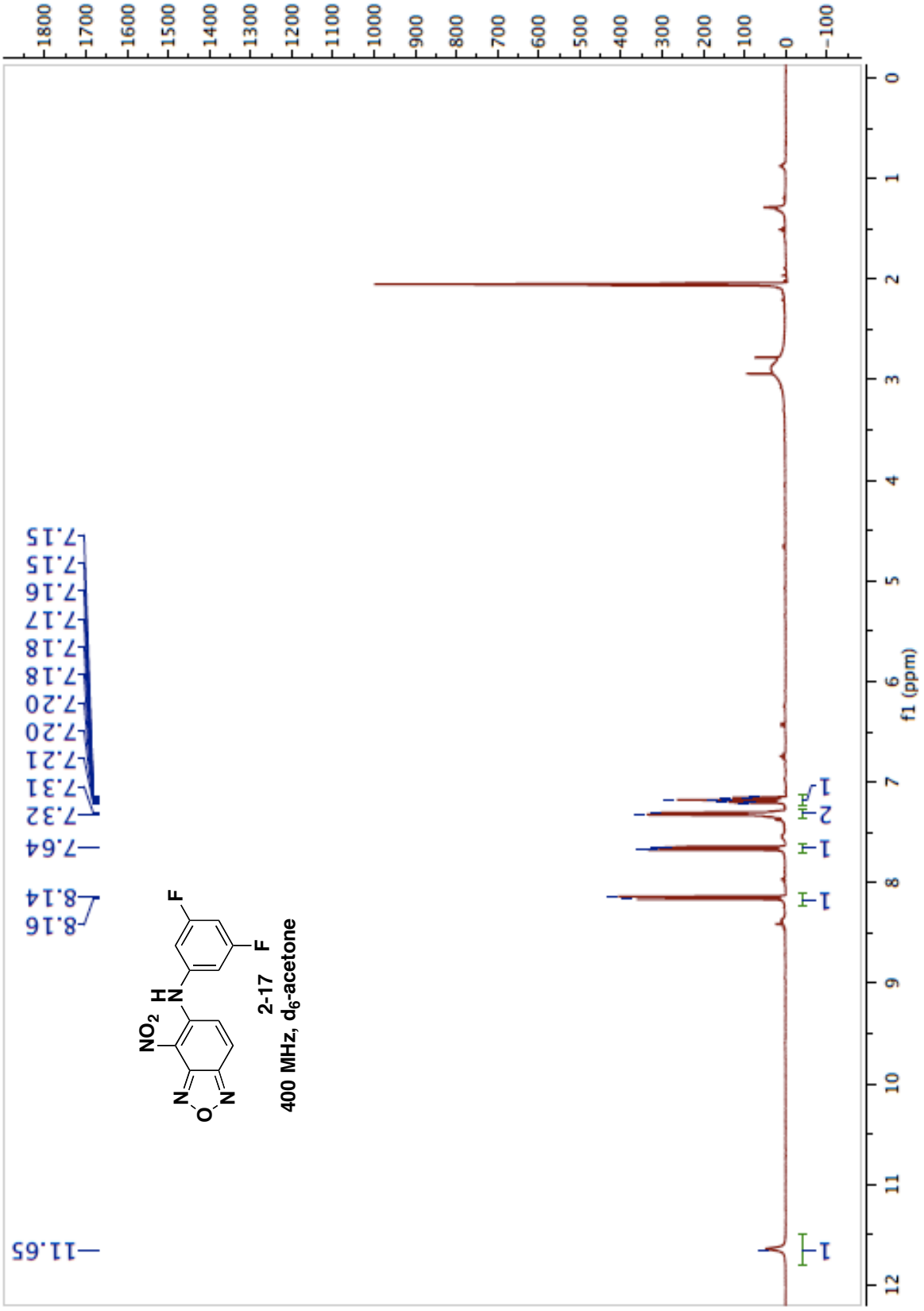

155.76
153.27
147.86
145.99
128.97
127.00
126.68
126.63
125.70
122.83
122.66
95.97
2-13
100 MHz, d6-acetone
f1 (ppm)

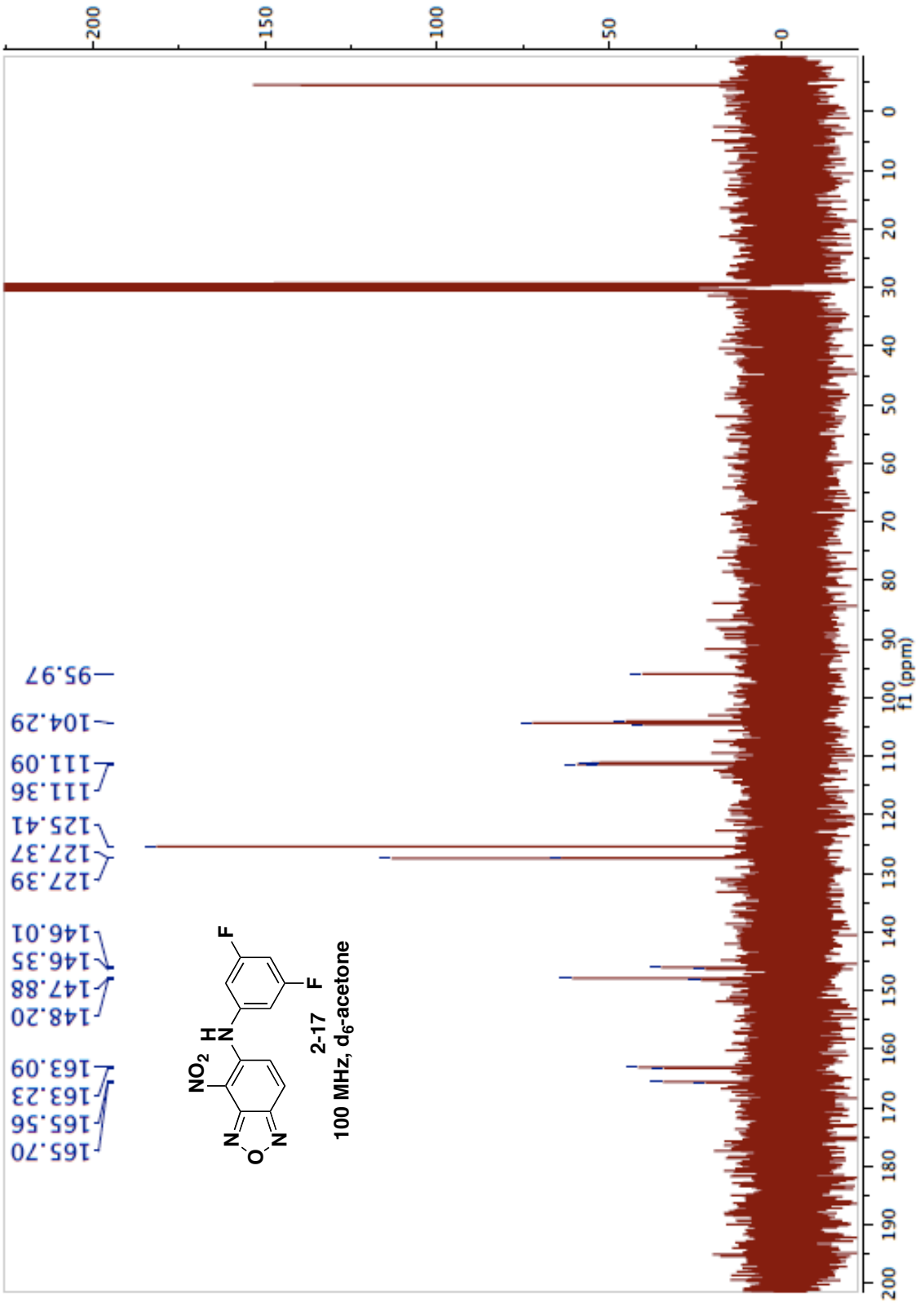

11.66
8.11
8.09
7.96
7.81
7.79
7.60
7.57
7.55
7.53
7.50
1
1
1
3
F
Cl
H
N
NO2
N
O
N
2-14
400 MHz, d6-acetone
f1 (ppm)

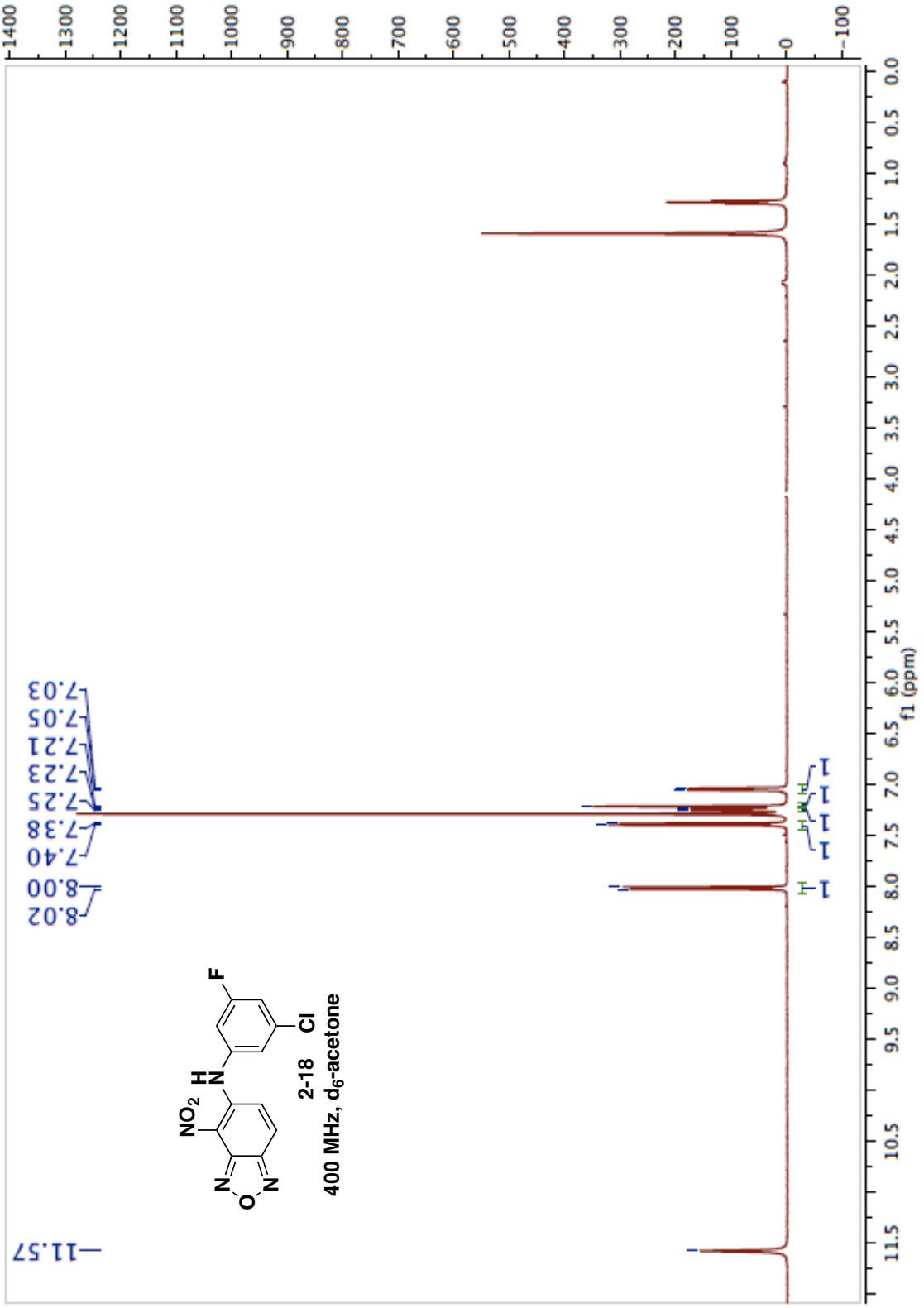

161.84
158.61
156.11
148.99
146.78
145.08
134.04
129.34
127.82
127.74
126.24
124.36
121.42
121.12
117.80
117.62
NO2
H
N
F
Cl
N
O
N
2-14
100 MHz, d6-acetone
f1 (ppm)

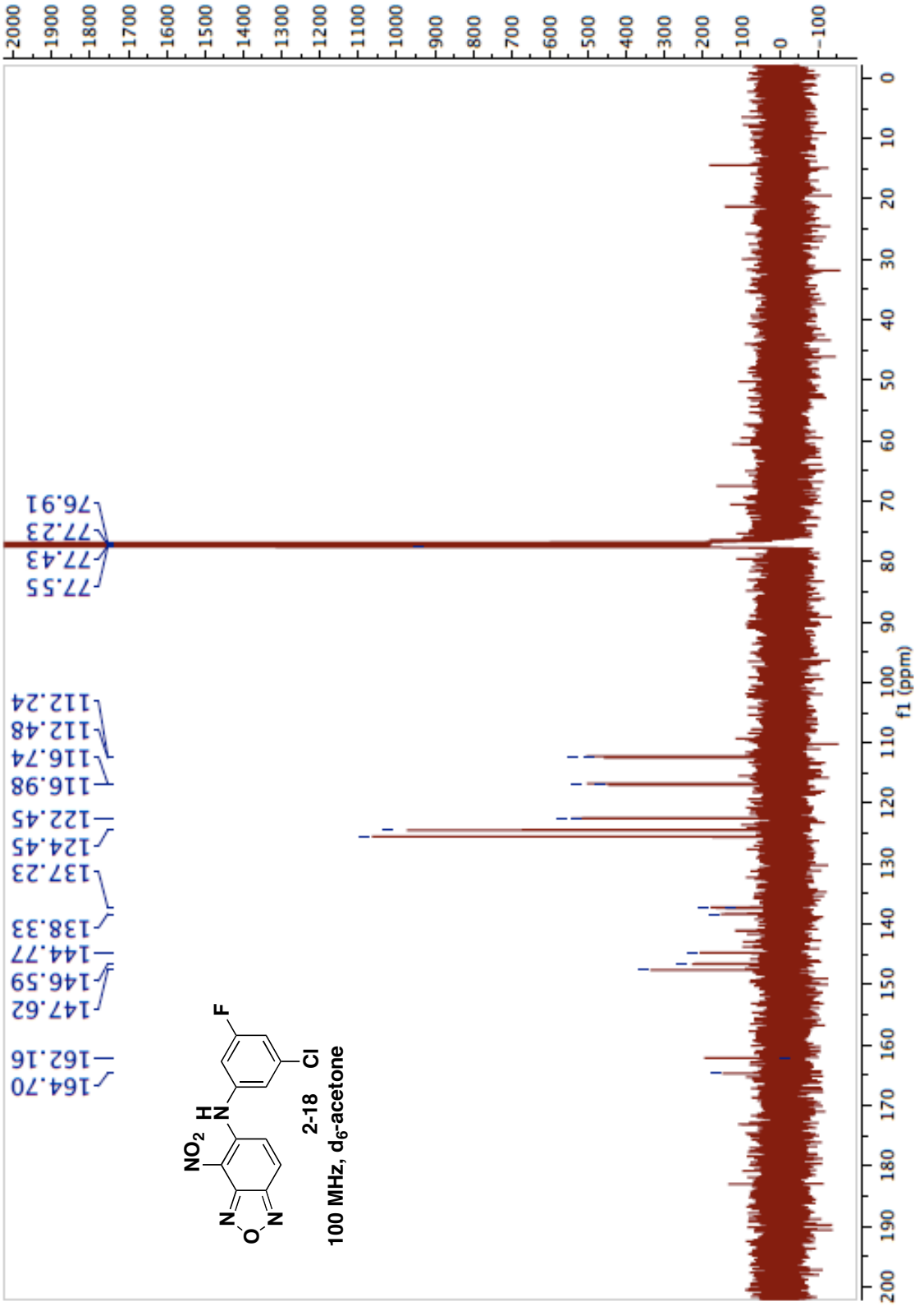

11.73
8.14
8.12
8.08
7.89
7.89
7.63
7.61
2-15
400 MHz, d6-acetone
f1 (ppm)

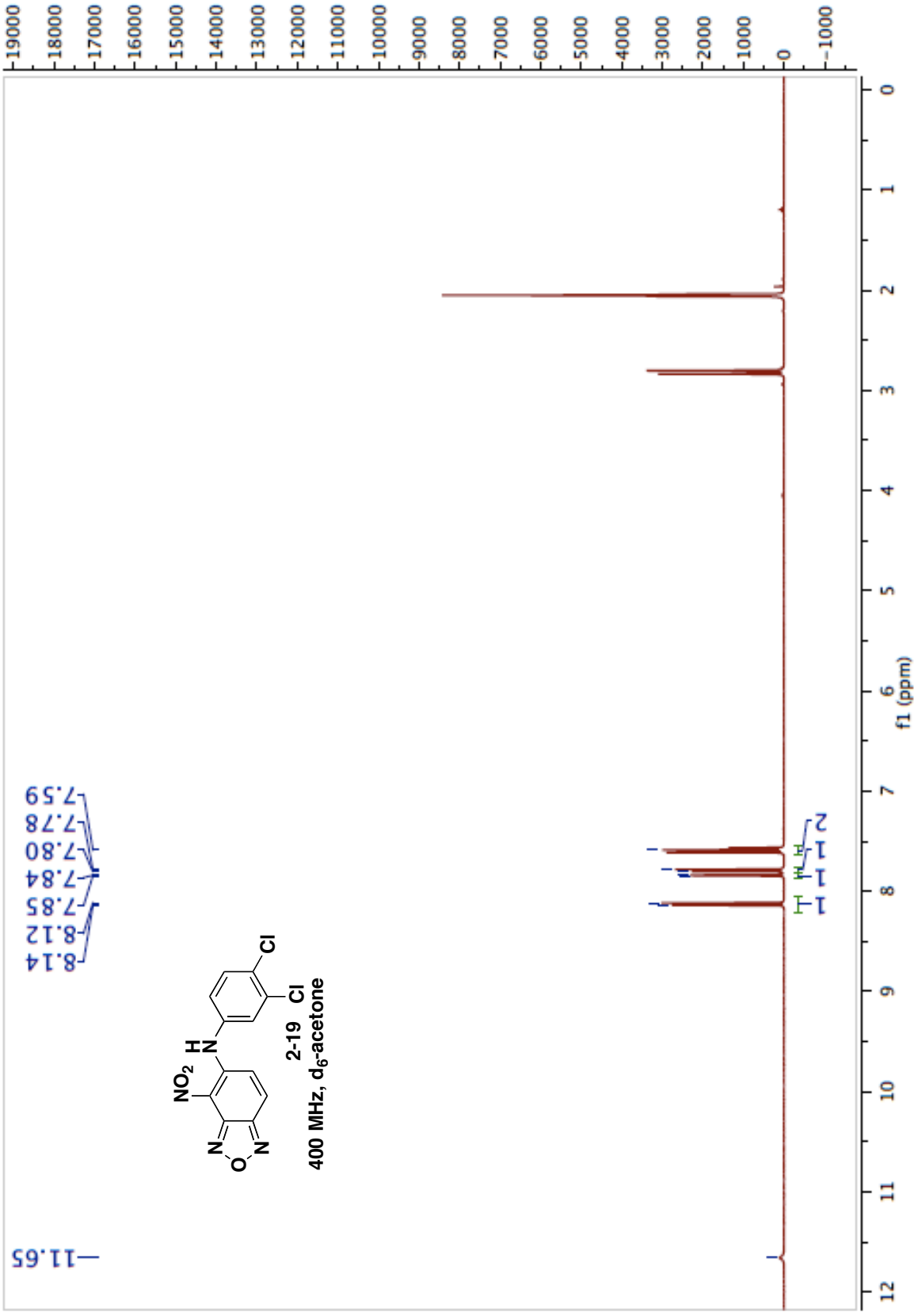

29.92
110.33
120.88
124.85
127.33
127.44
129.38
132.96
146.04
147.83
147.88
f1 (ppm)
2-15
100 MHz, d6-acetone

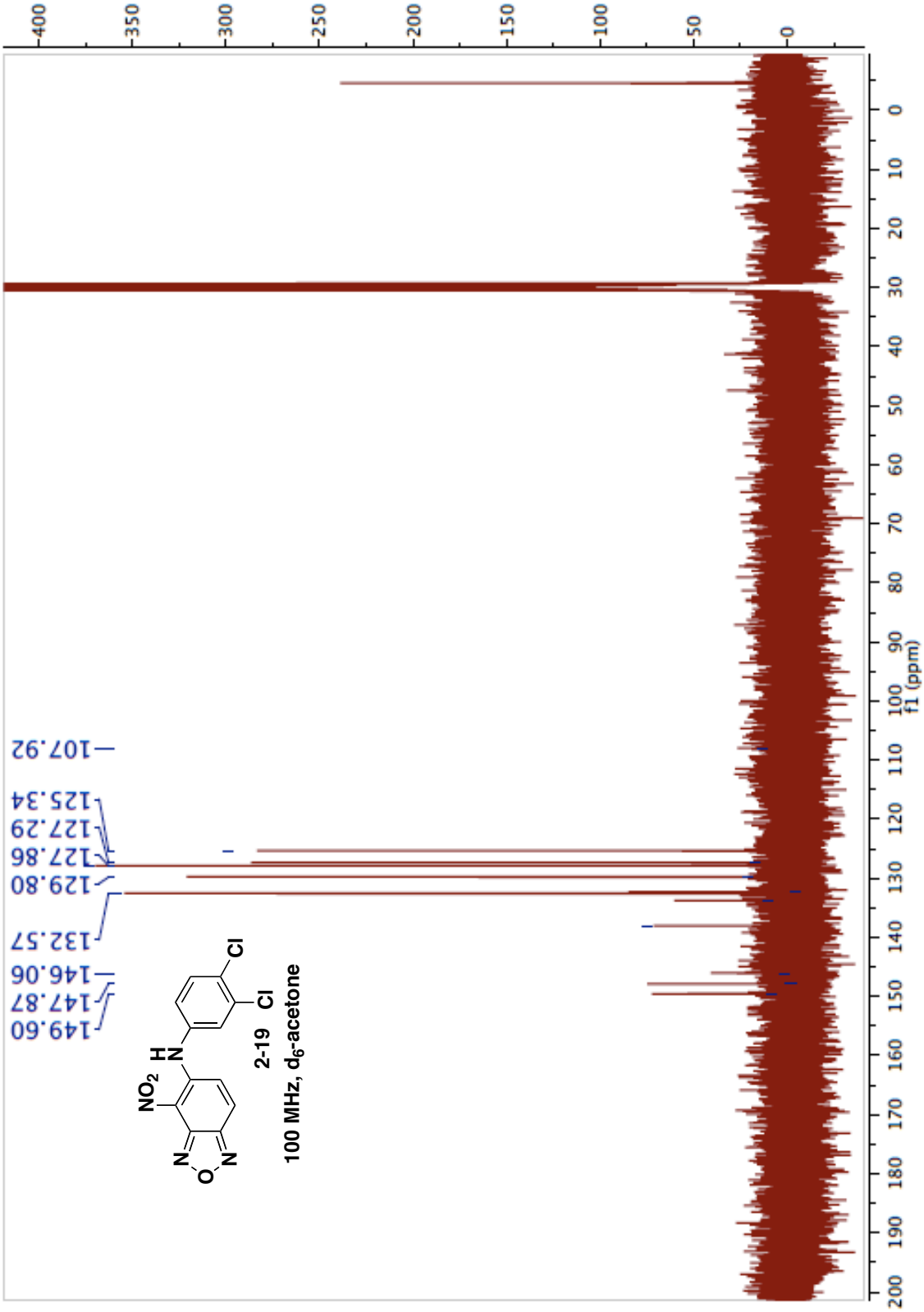

11.65
8.16
8.14
7.64
7.32
7.31
7.21
7.20
7.20
7.18
7.18
7.17
7.16
7.15
7.15
1
1
2
1
1
1
2-17
400 MHz, d6-acetone
f1 (ppm)

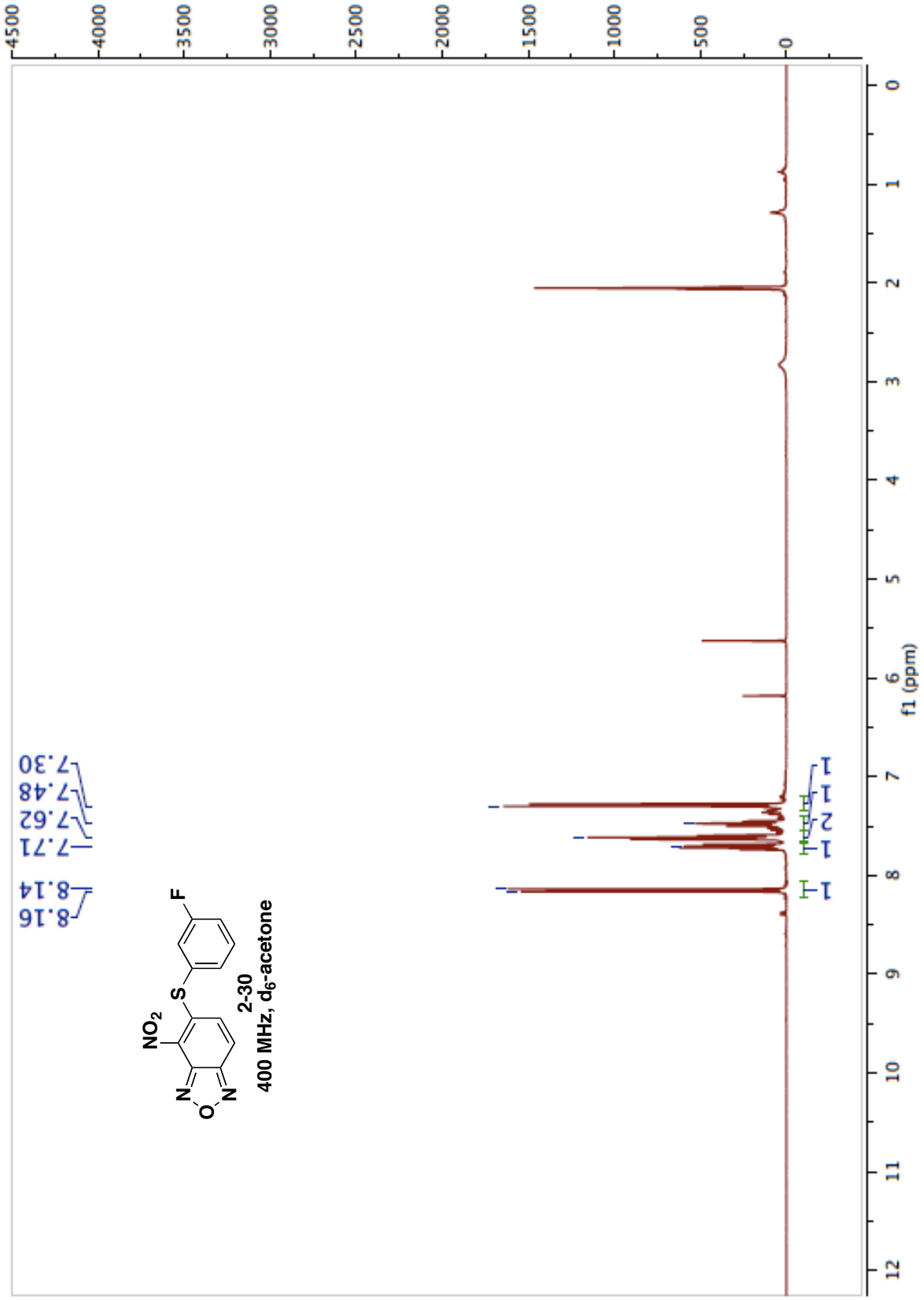

95.97
104.29
111.09
111.36
125.41
127.37
127.39
146.01
146.35
147.88
148.20
163.09
163.23
165.56
165.70
F
H
N
NO2
N
O
N
F
2-17
100 MHz, d6-acetone
f1 (ppm)

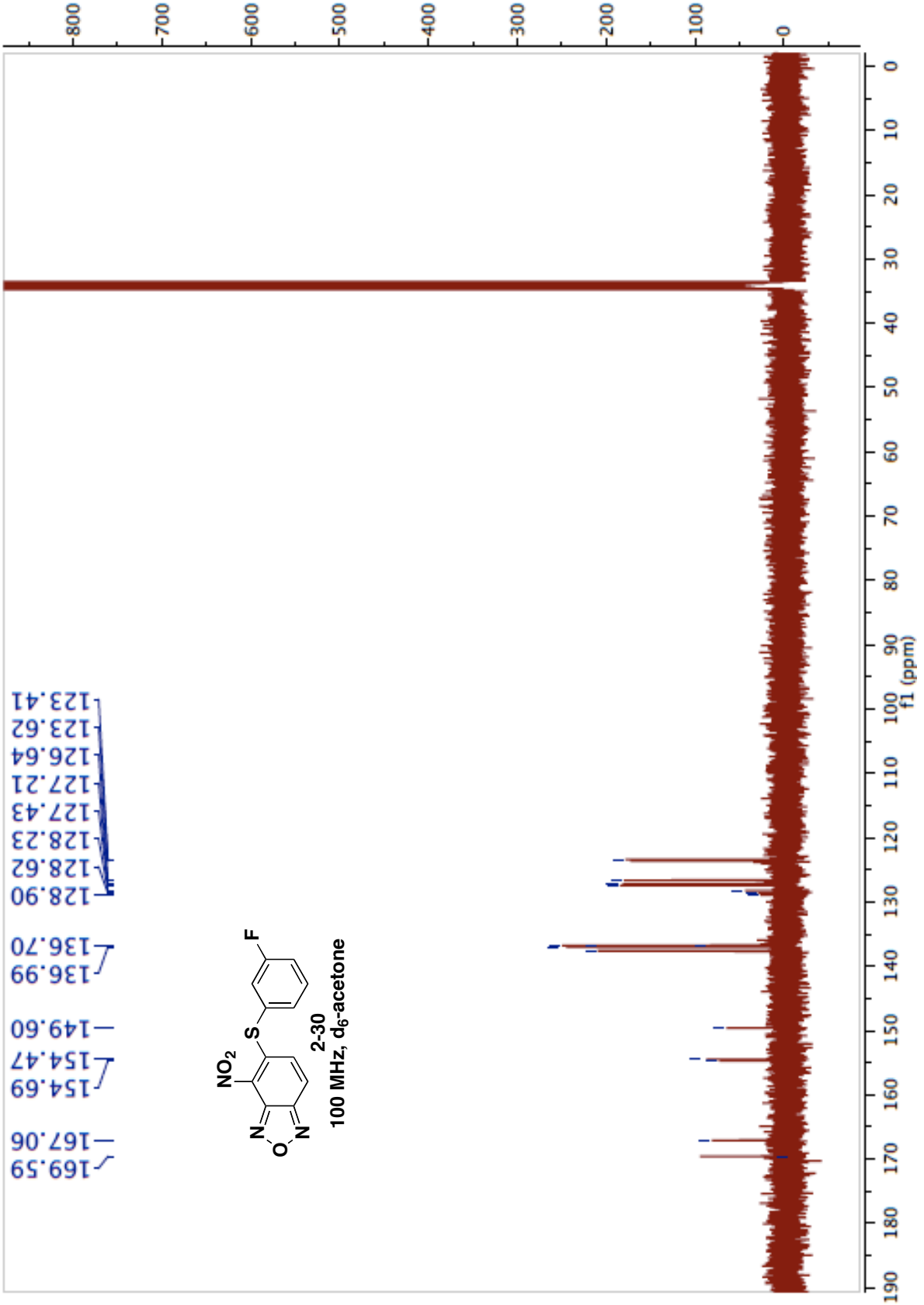

11.57
8.02
8.00
7.40
7.38
7.25
7.23
7.21
7.05
7.03
2-18
400 MHz, d6-acetone
f1 (ppm)

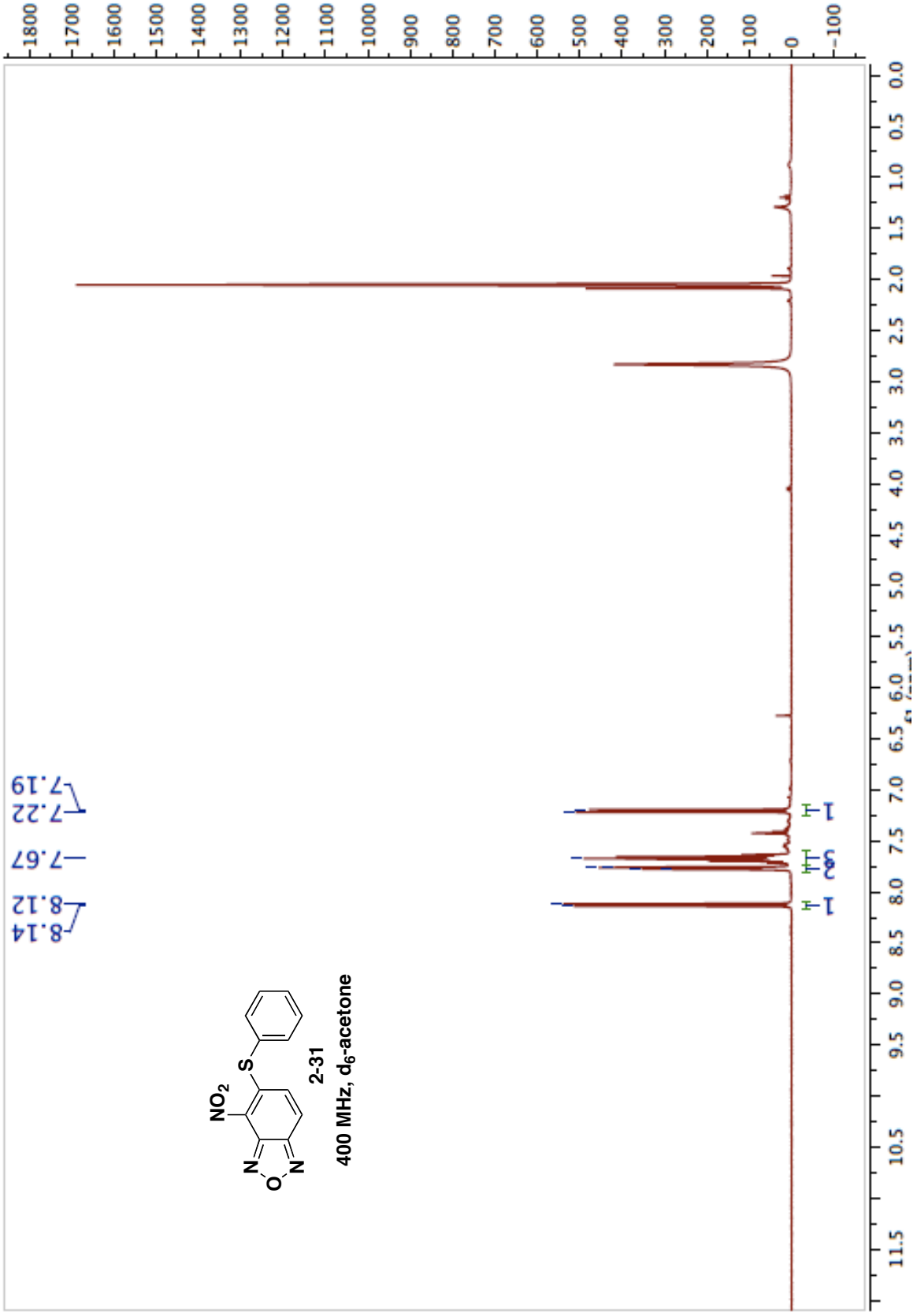

77.55
77.43
77.23
76.91
116.98
116.74
112.48
112.24
122.45
124.45
137.23
138.33
144.77
146.59
147.62
162.16
164.70
2-18
100 MHz, d6-acetone
f1 (ppm)

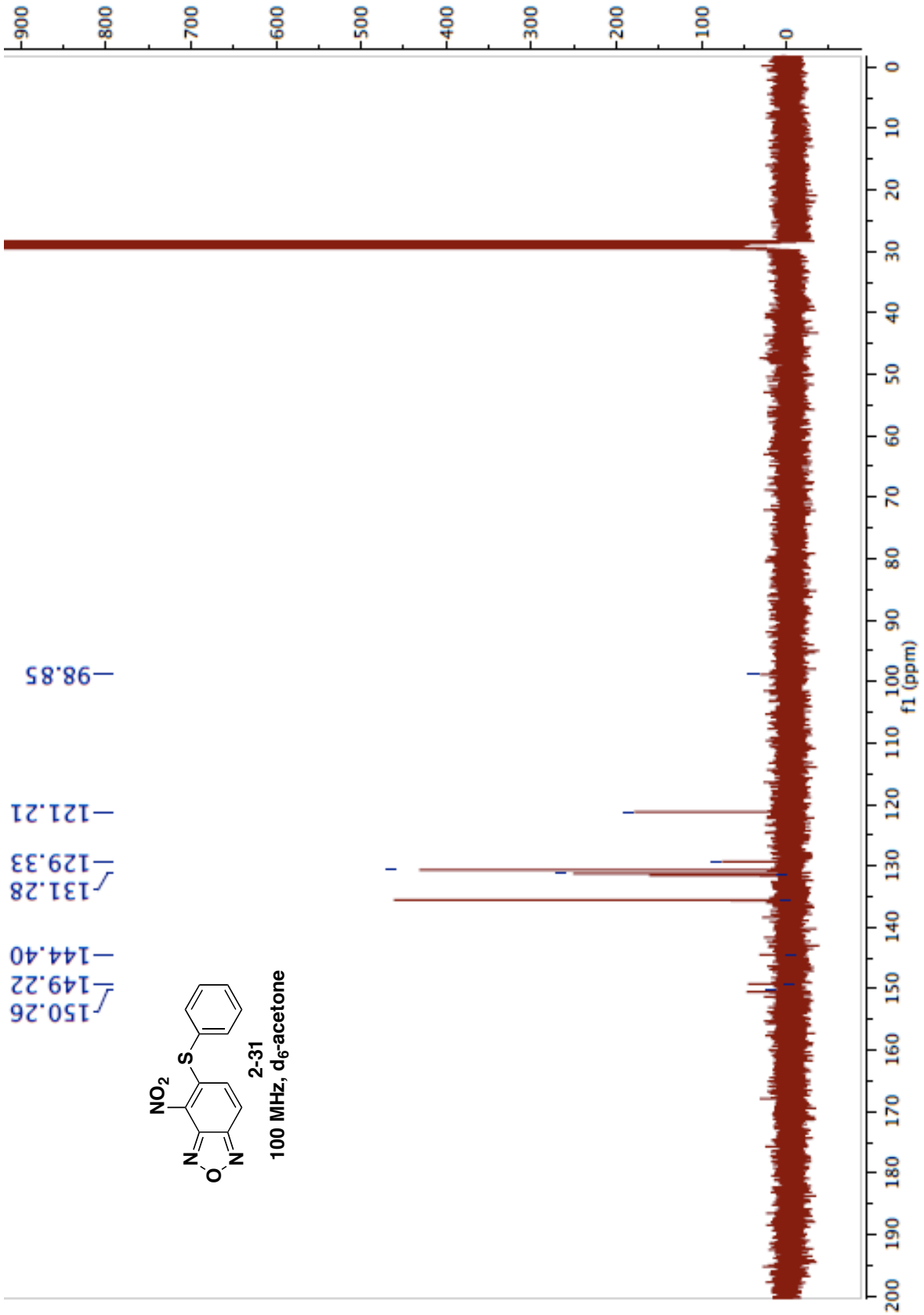

11.65
8.14
8.12
7.85
7.84
7.80
7.78
7.59
1
1
1
2
f1 (ppm)
2-19
400 MHz, d6-acetone

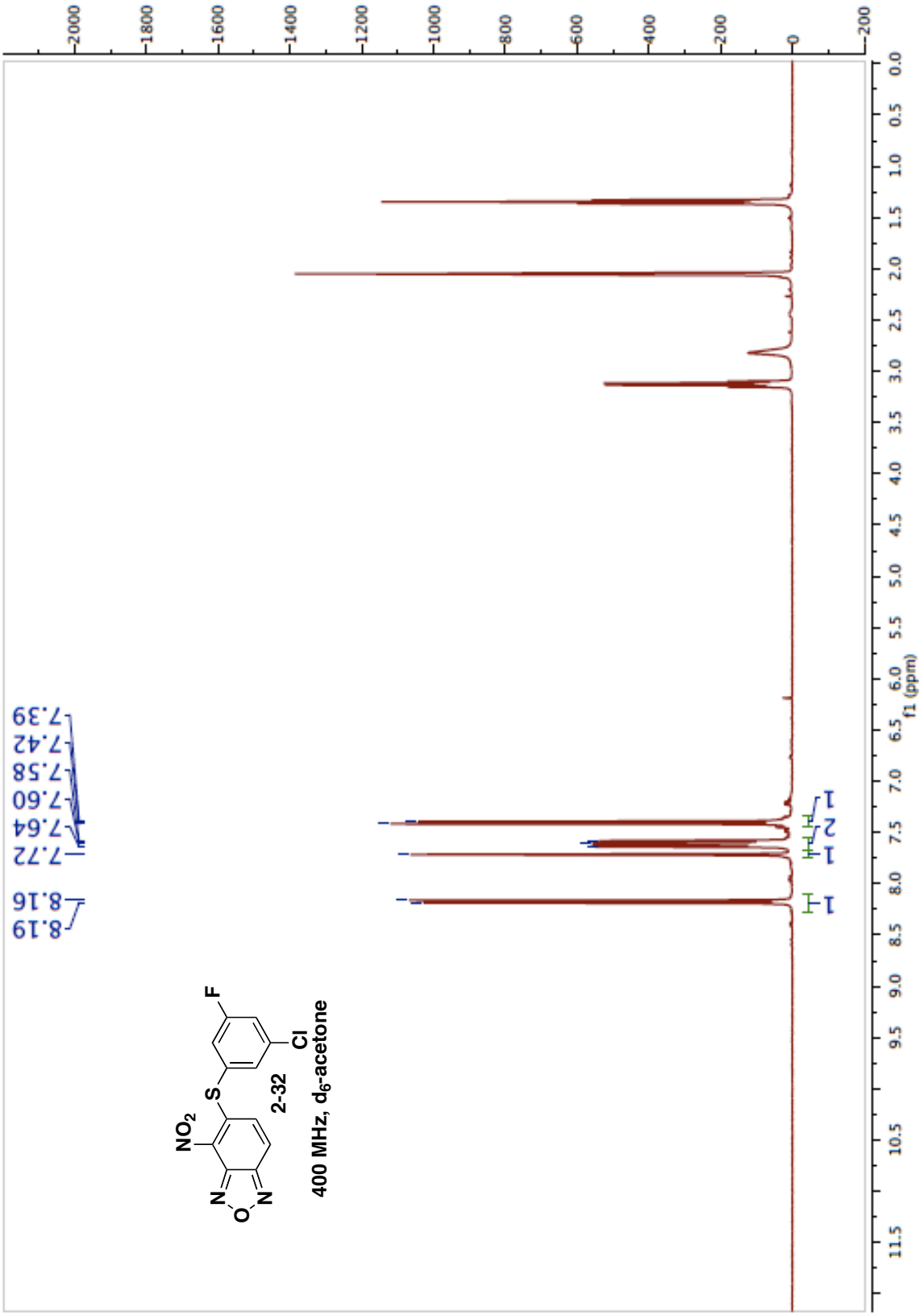

107.92
125.34
127.29
127.86
129.80
132.57
146.06
147.87
149.60
Cl
Cl
H
N
NO2
N
O
N
2-19
100 MHz, d6-acetone
f1 (ppm)

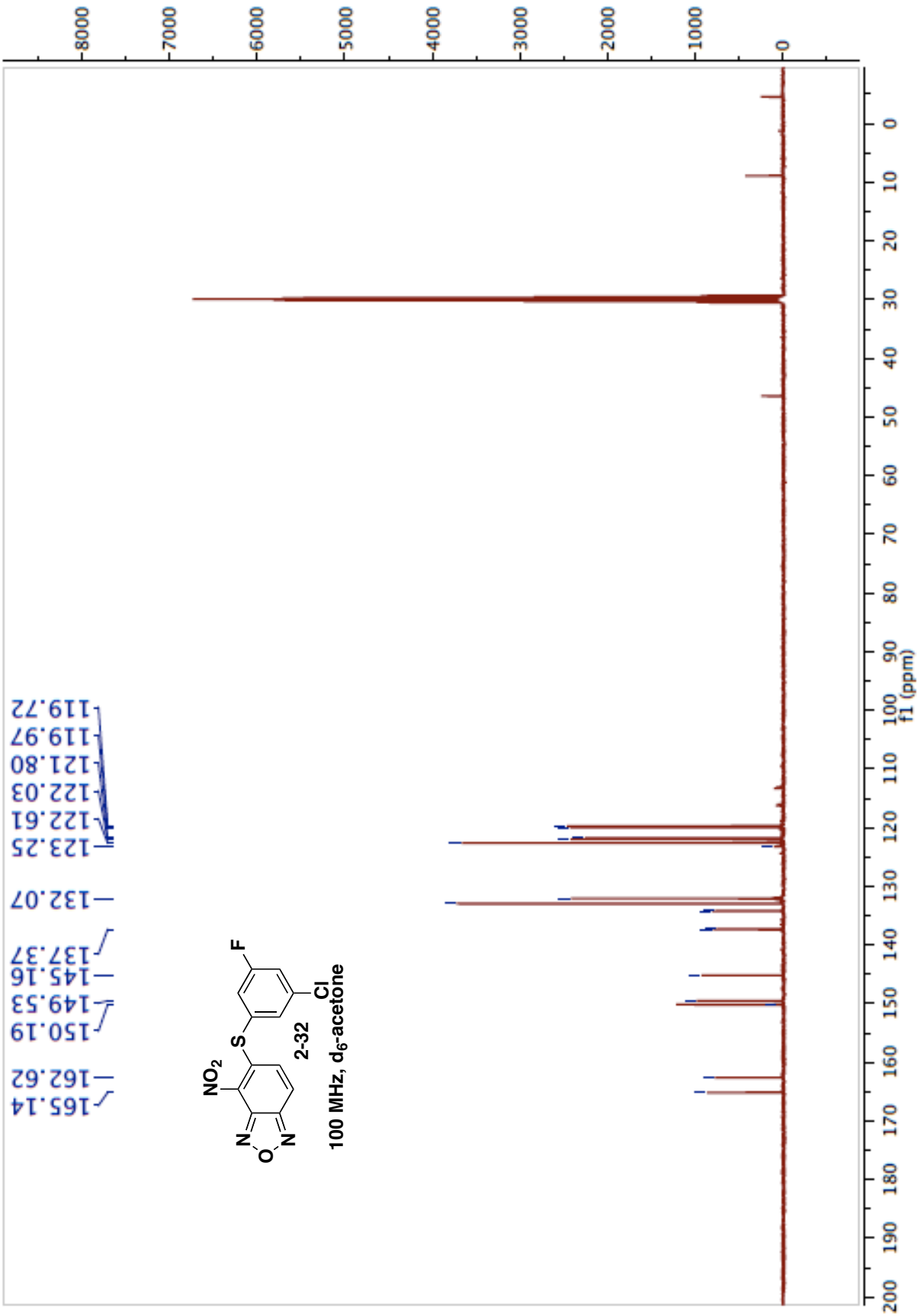

8.16
8.14
7.71
7.62
7.48
7.30
F
S
NO2
N
O
N
2-30
400 MHz, d6-acetone
f1 (ppm)

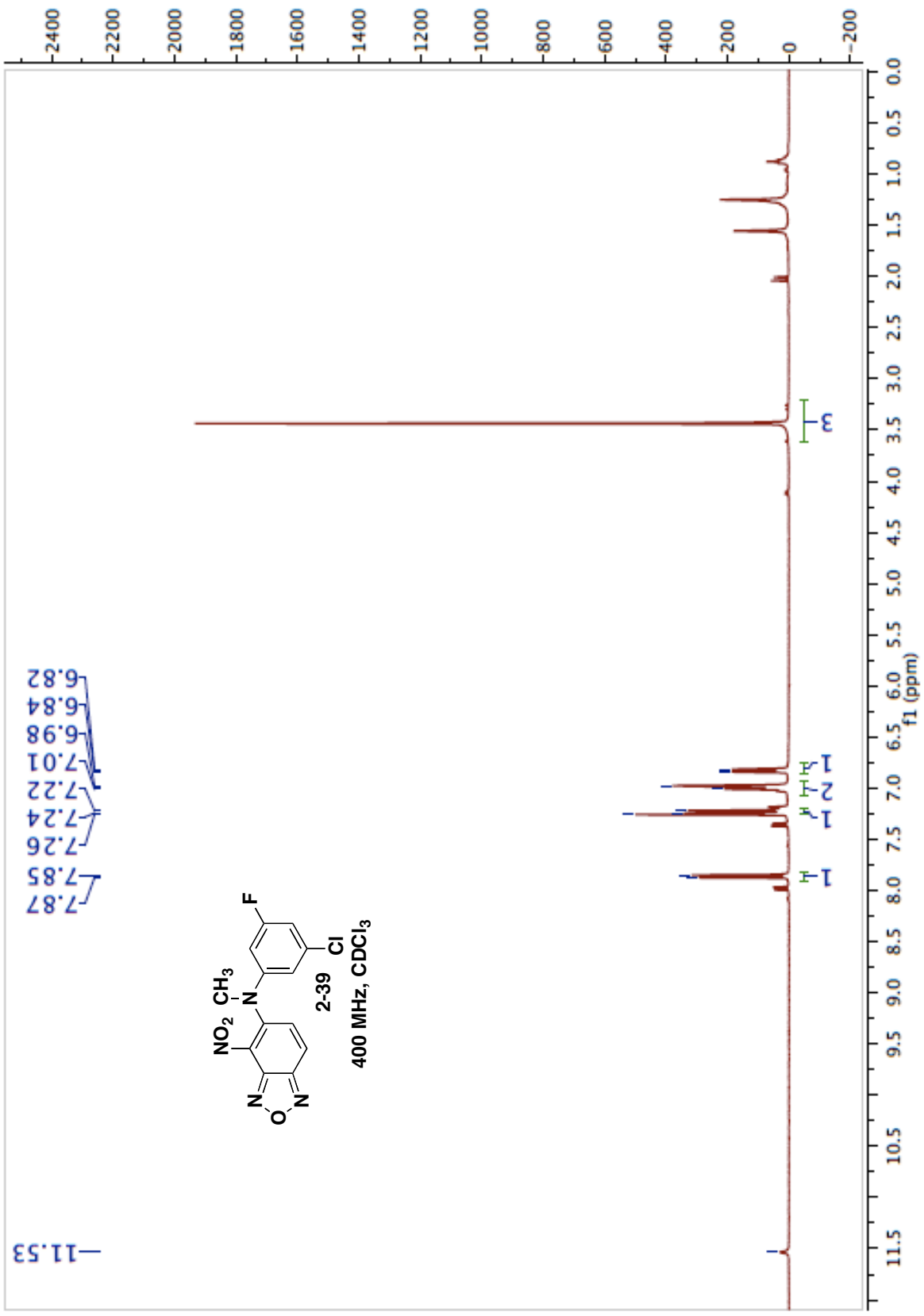

123.41
123.62
126.64
127.21
127.43
128.23
128.62
128.90
136.70
136.99
149.60
154.47
154.69
167.06
169.59
F
S
NO2
N
O
N
2-30
100 MHz, d6-acetone
f1 (ppm)

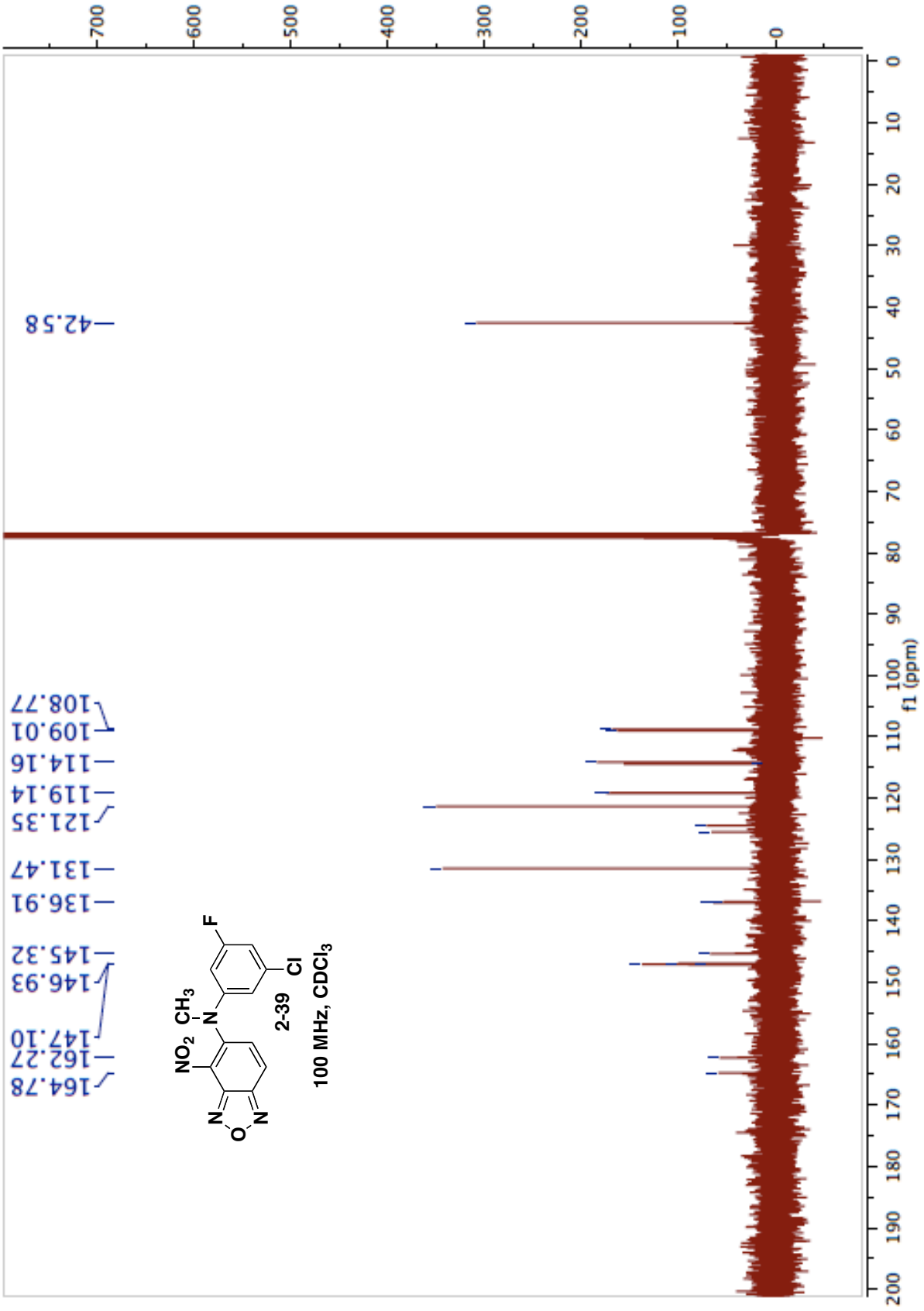

8.14
8.12
7.67
7.22
7.19
1H
3H
1H
2-31
400 MHz, d6-acetone

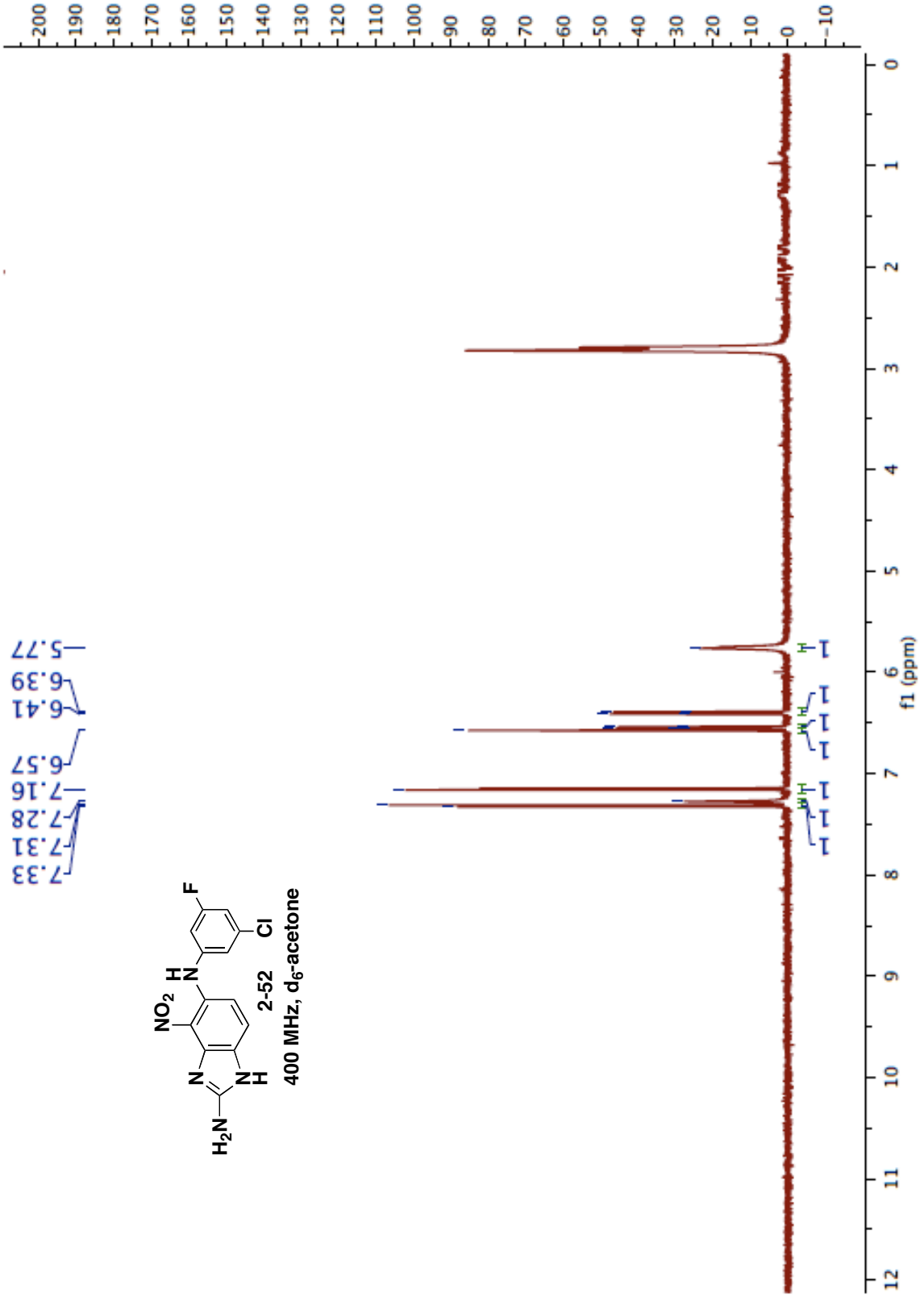

150.26
149.22
144.40
131.28
129.33
121.21
98.85
2-31
100 MHz, d6-acetone
f1 (ppm)

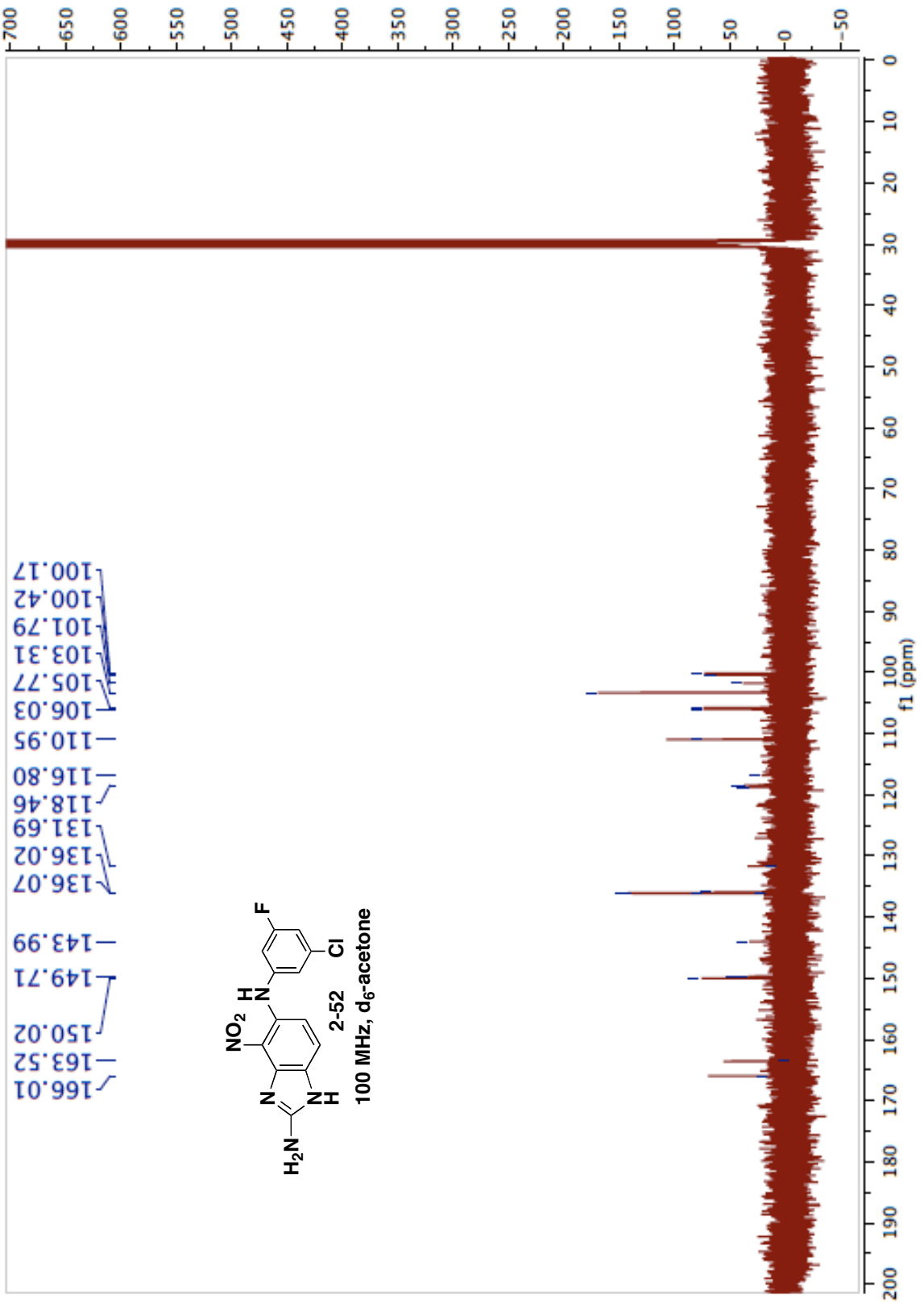

8.19
8.16
7.72
7.64
7.60
7.58
7.42
7.39
1
2
1
1
F
Cl
S
NO2
N
O
N
2-32
400 MHz, d6-acetone
f1 (ppm)

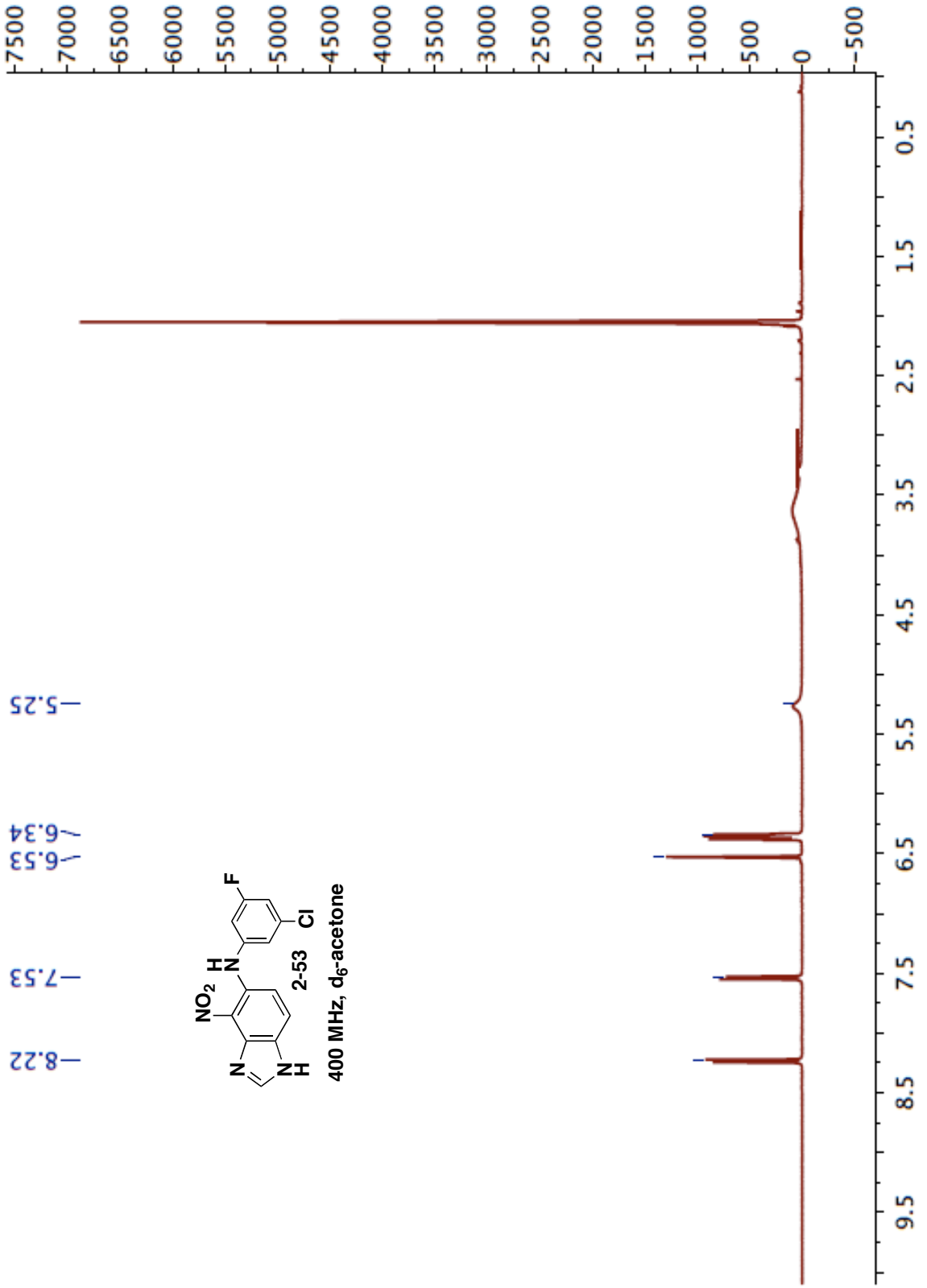

8000
7000
6000
5000
4000
3000
2000
1000
0
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
f1 (ppm)
119.72
119.97
121.80
122.03
122.61
123.25
132.07
137.37
145.16
149.53
150.19
162.62
165.14
F
S
Cl
NO2
N
O
N
2-32
100 MHz, d6-acetone

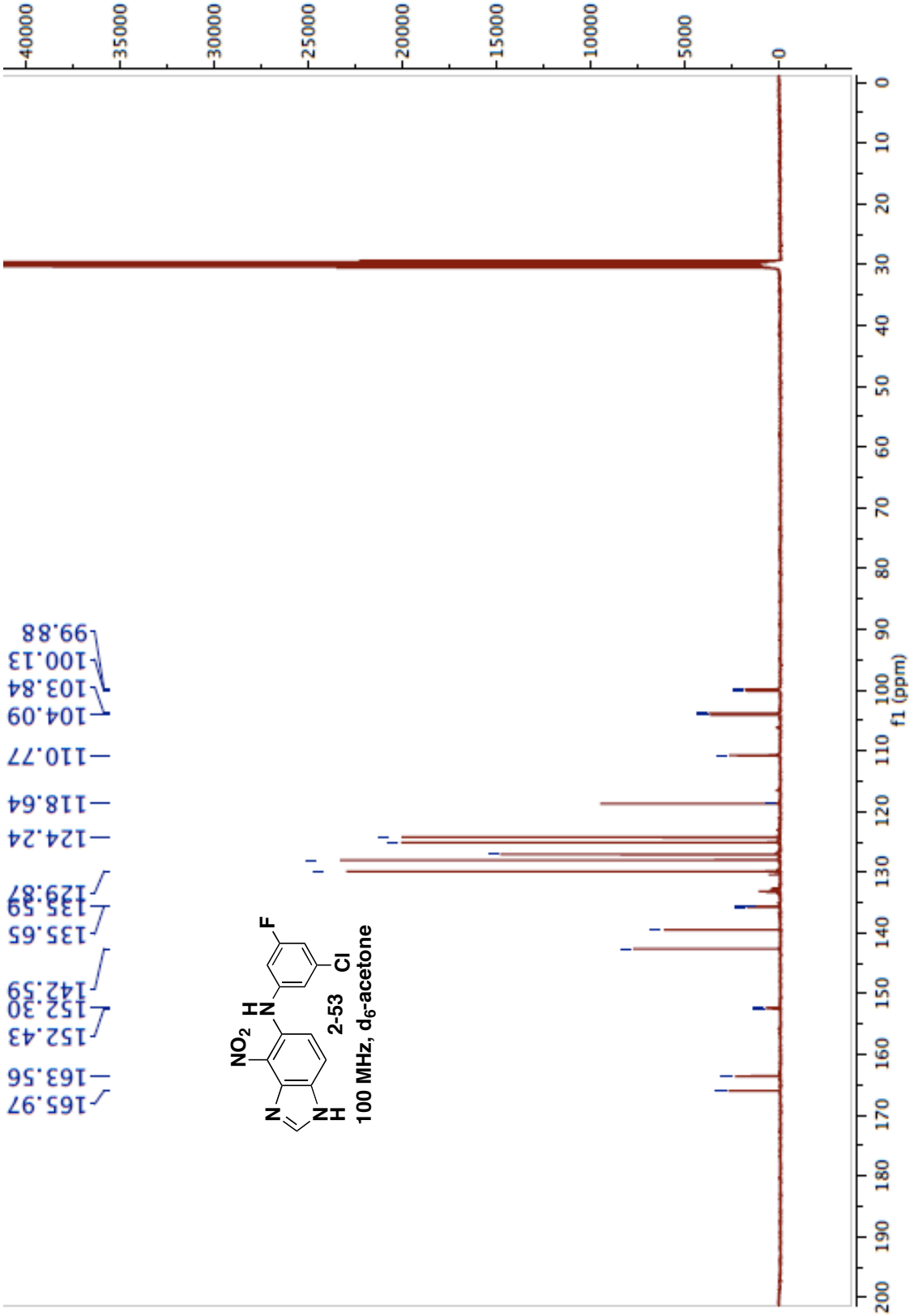

11.53
7.87
7.85
7.26
7.24
7.22
7.01
6.98
6.84
6.82
3
1
1
2
1
1
f1 (ppm)
2-39
400 MHz, CDCl3

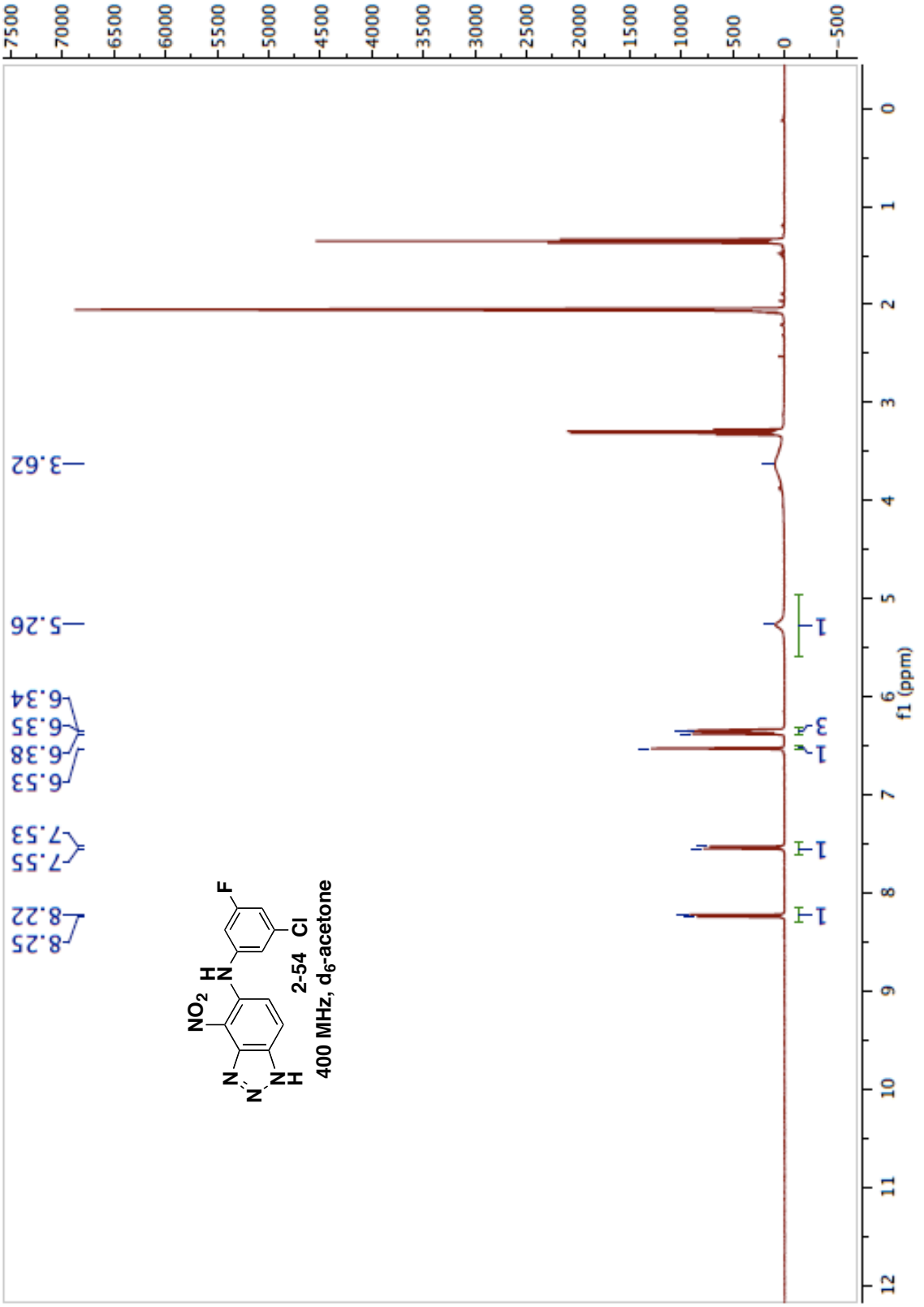

42.58
108.77
109.01
114.16
119.14
121.35
131.47
136.91
145.32
146.93
147.10
162.27
164.78
NO2 CH3 N
F
Cl
2-39
100 MHz, CDCl3
f1 (ppm)

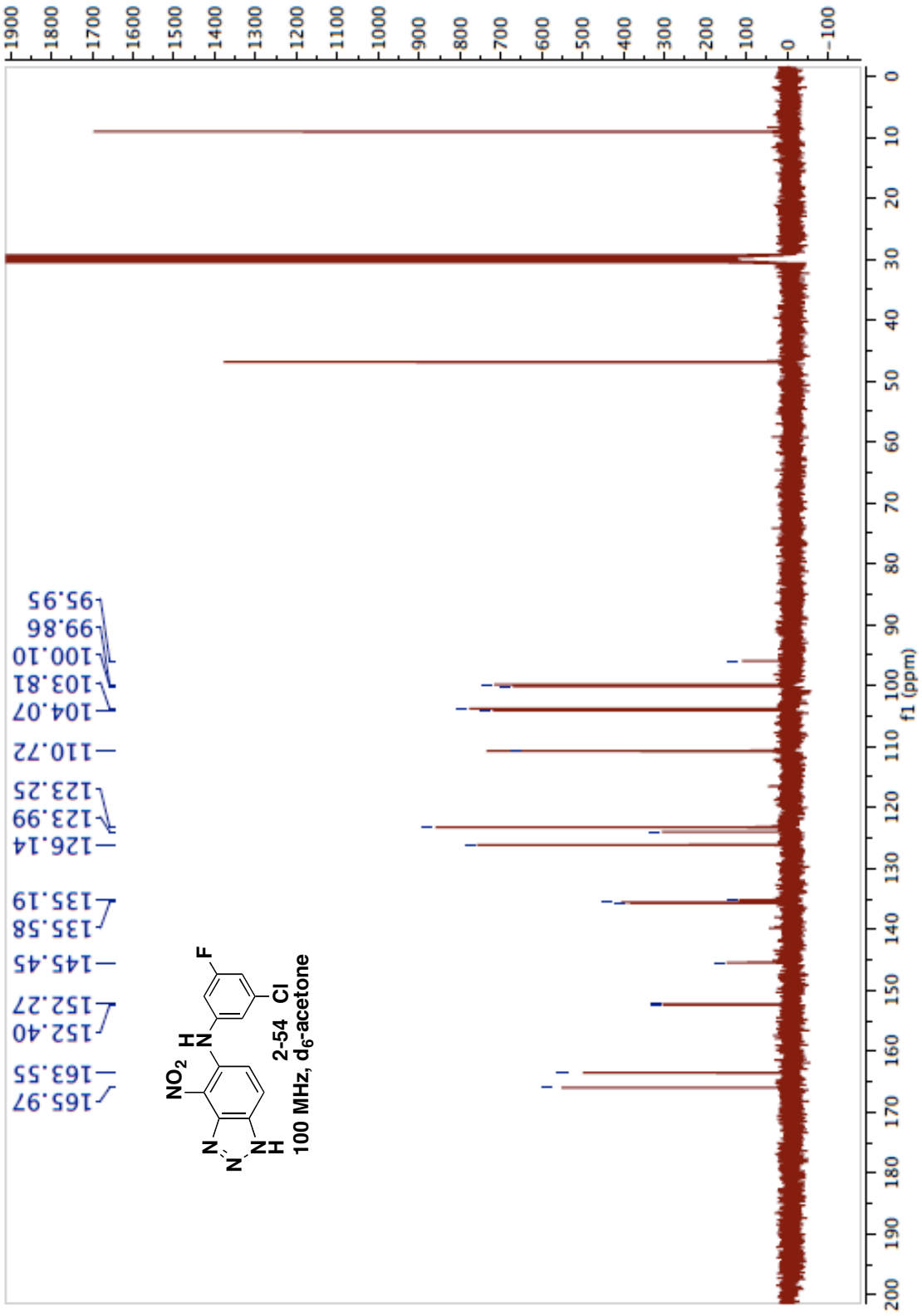

7.33
7.31
7.28
7.16
6.57
6.41
6.39
5.77
2-52
400 MHz, d6-acetone
f1 (ppm)

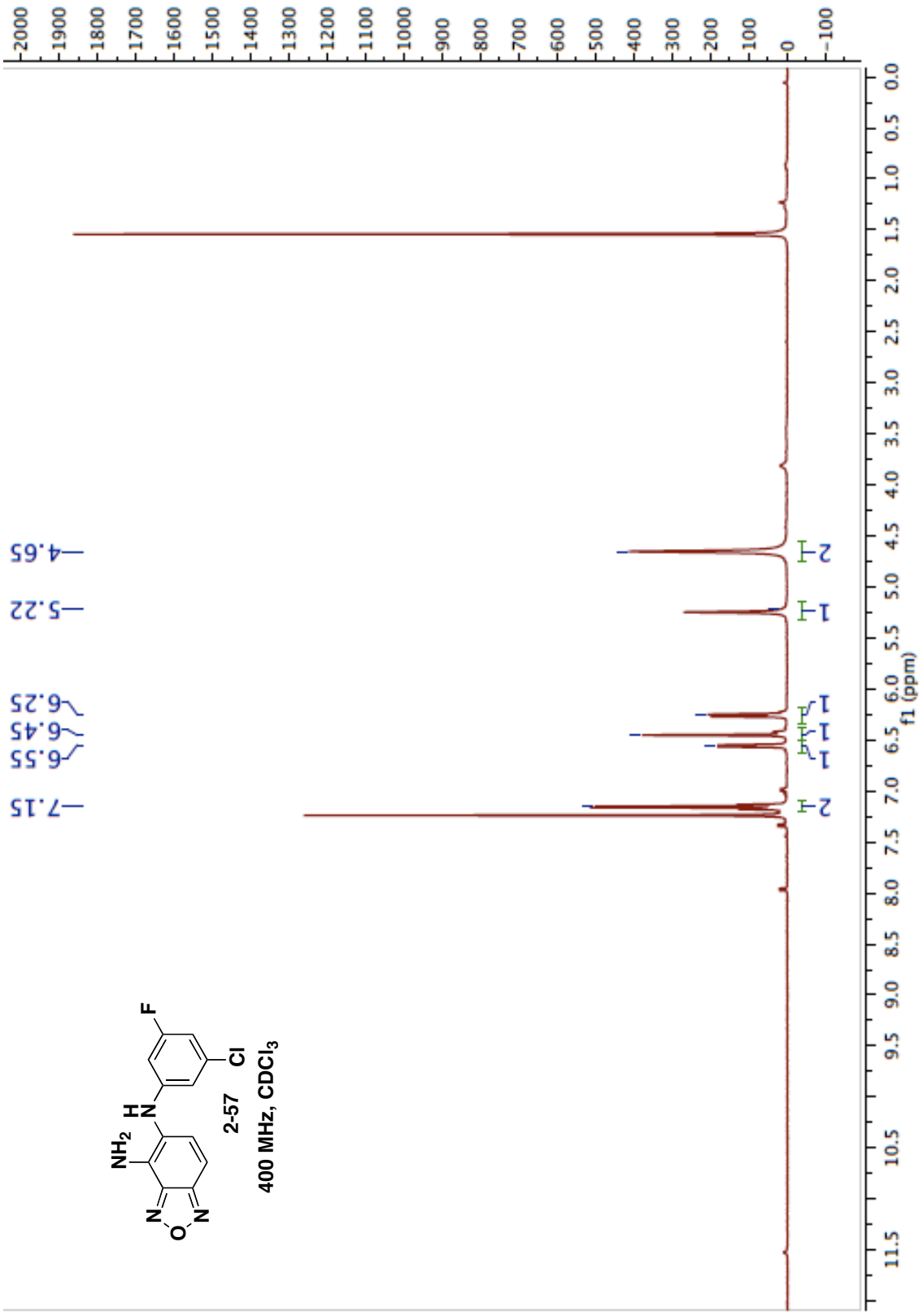

100.17
100.42
101.79
103.31
105.77
106.03
110.95
116.80
118.46
131.69
136.02
136.07
143.99
149.71
150.02
163.52
166.01
2-52
100 MHz, d6-acetone
f1 (ppm)

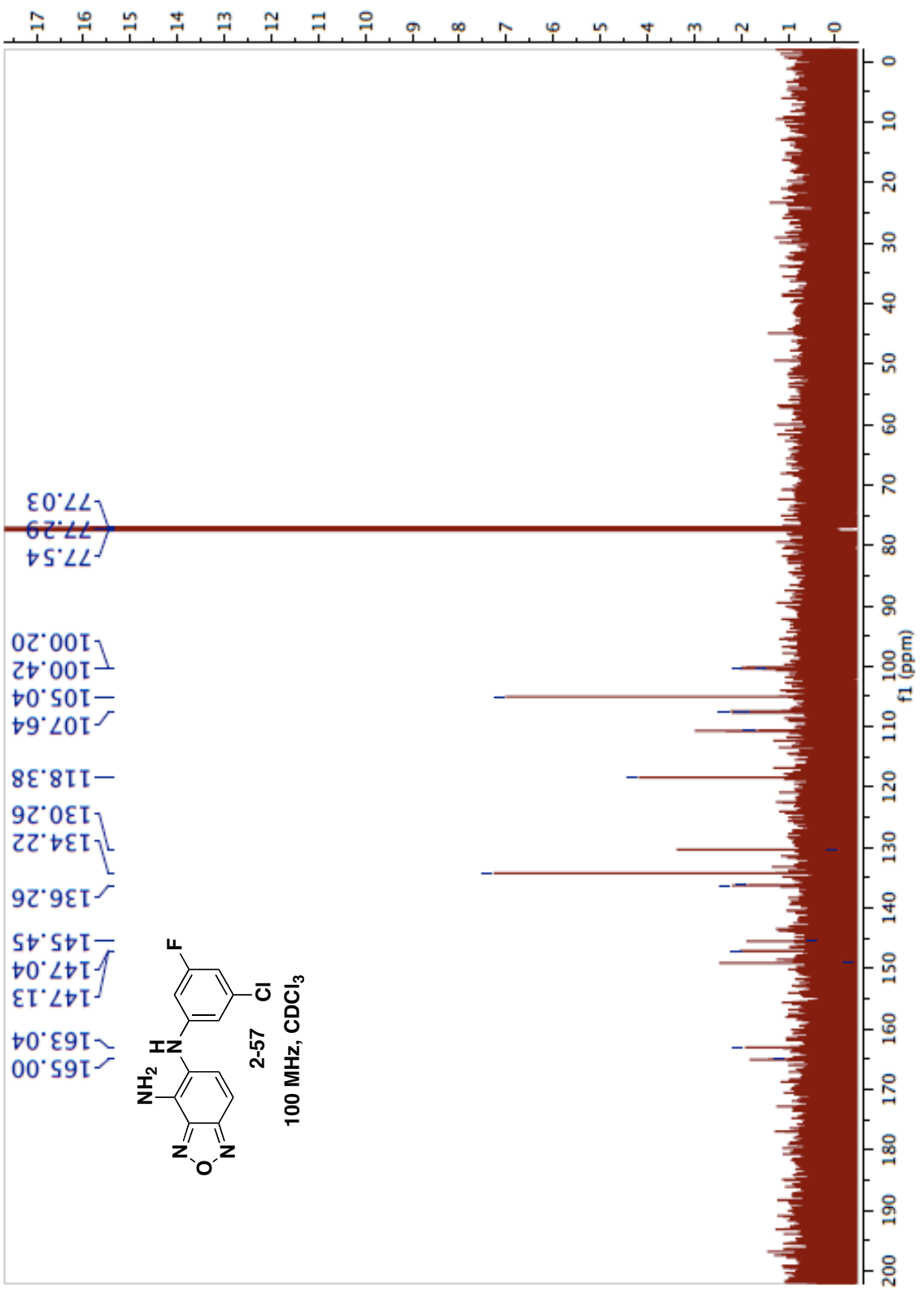

8.22
7.53
6.53
6.34
5.25
2-53
400 MHz, d6-acetone

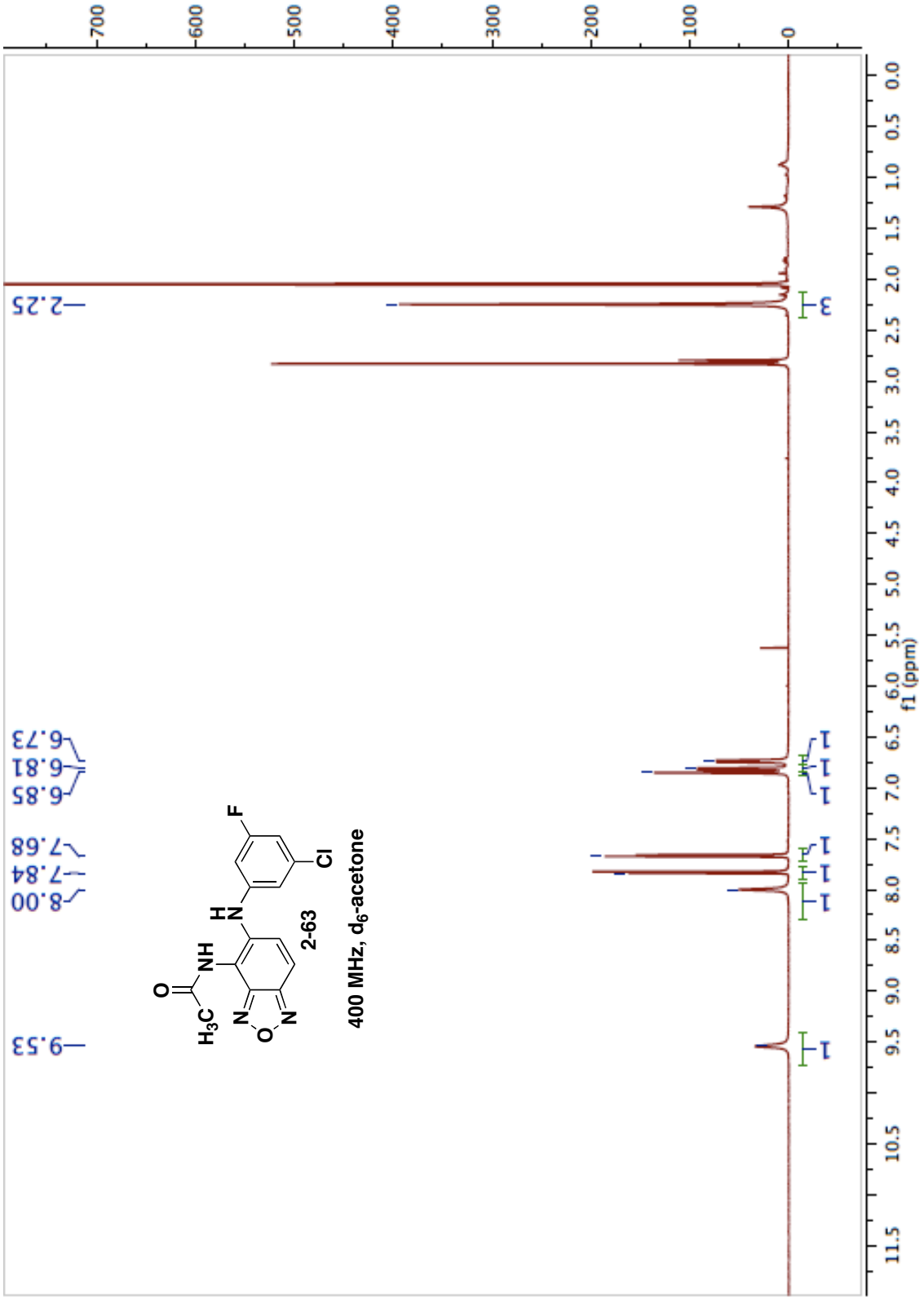

99.88
100.13
103.84
104.09
110.77
118.64
124.24
129.87
135.59
135.65
142.59
152.30
152.43
163.56
165.97
2-53
100 MHz, d6-acetone
f1 (ppm)

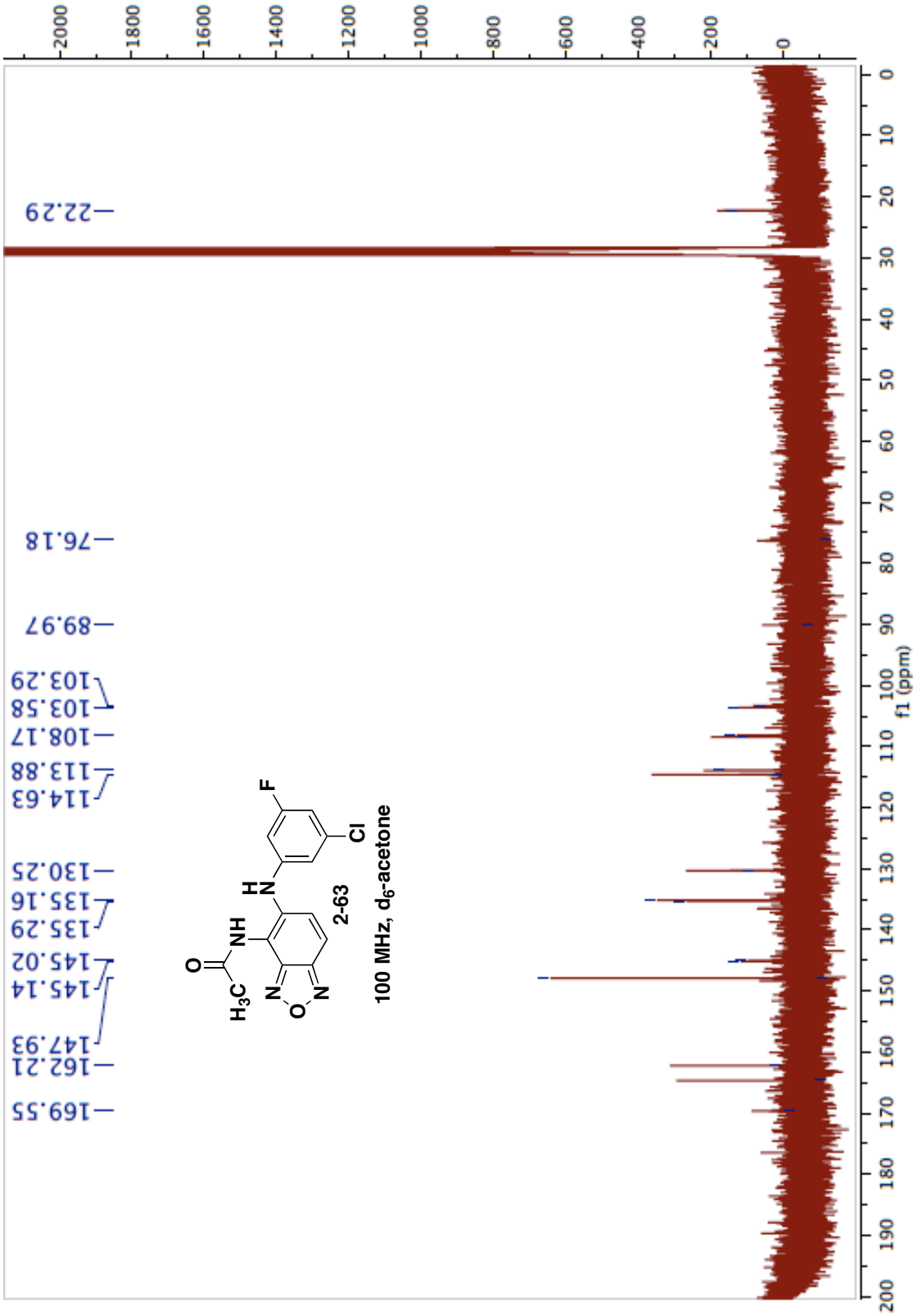

2-54 Cl
400 MHz, d6-acetone
f1 (ppm)

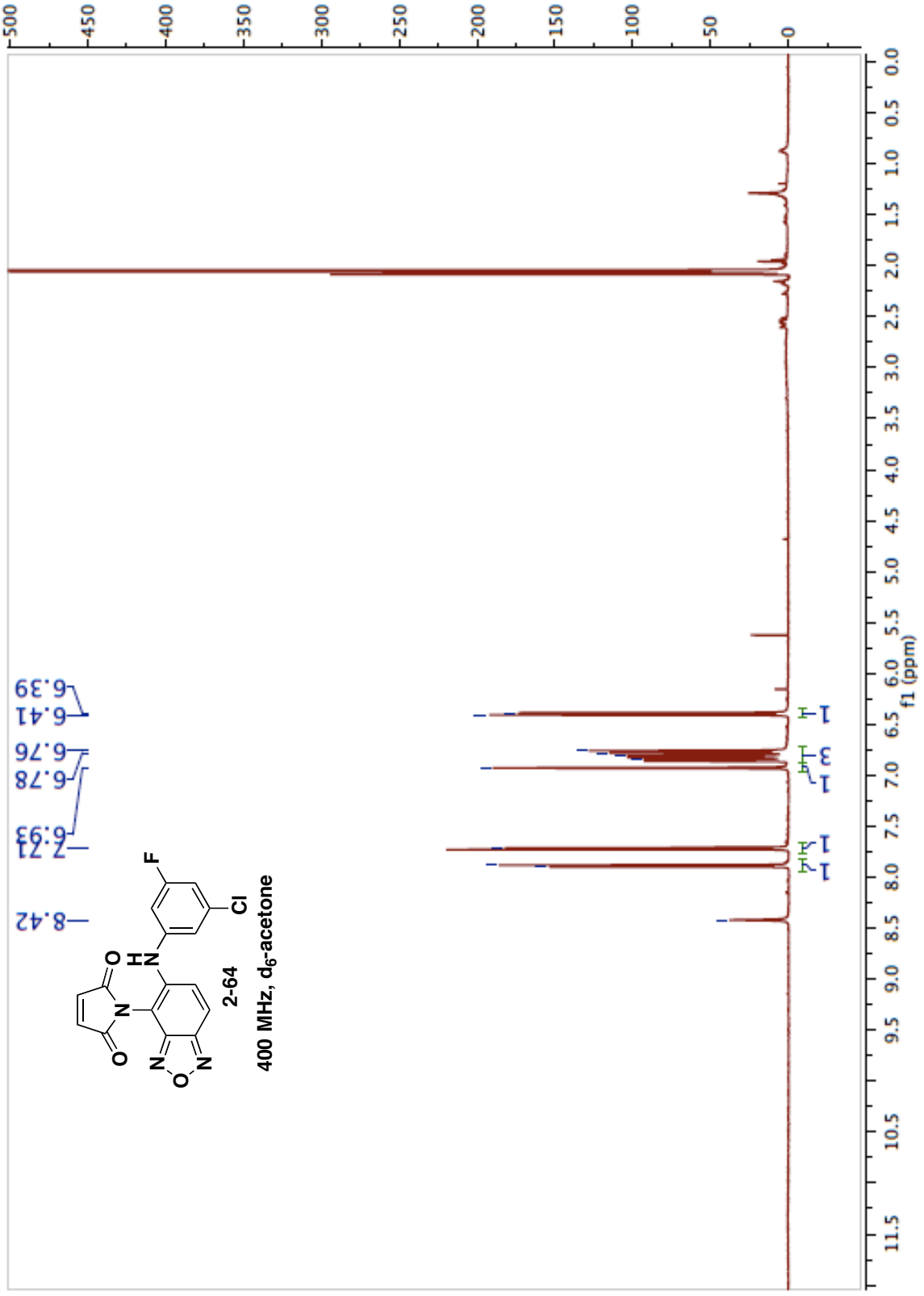

95.95
99.86
100.10
103.81
104.07
110.72
123.25
123.99
126.14
135.19
135.58
145.45
152.27
152.40
163.55
165.97
NO2
H
N
F
N
N
H
2-54 Cl
100 MHz, d6-acetone
f1 (ppm)

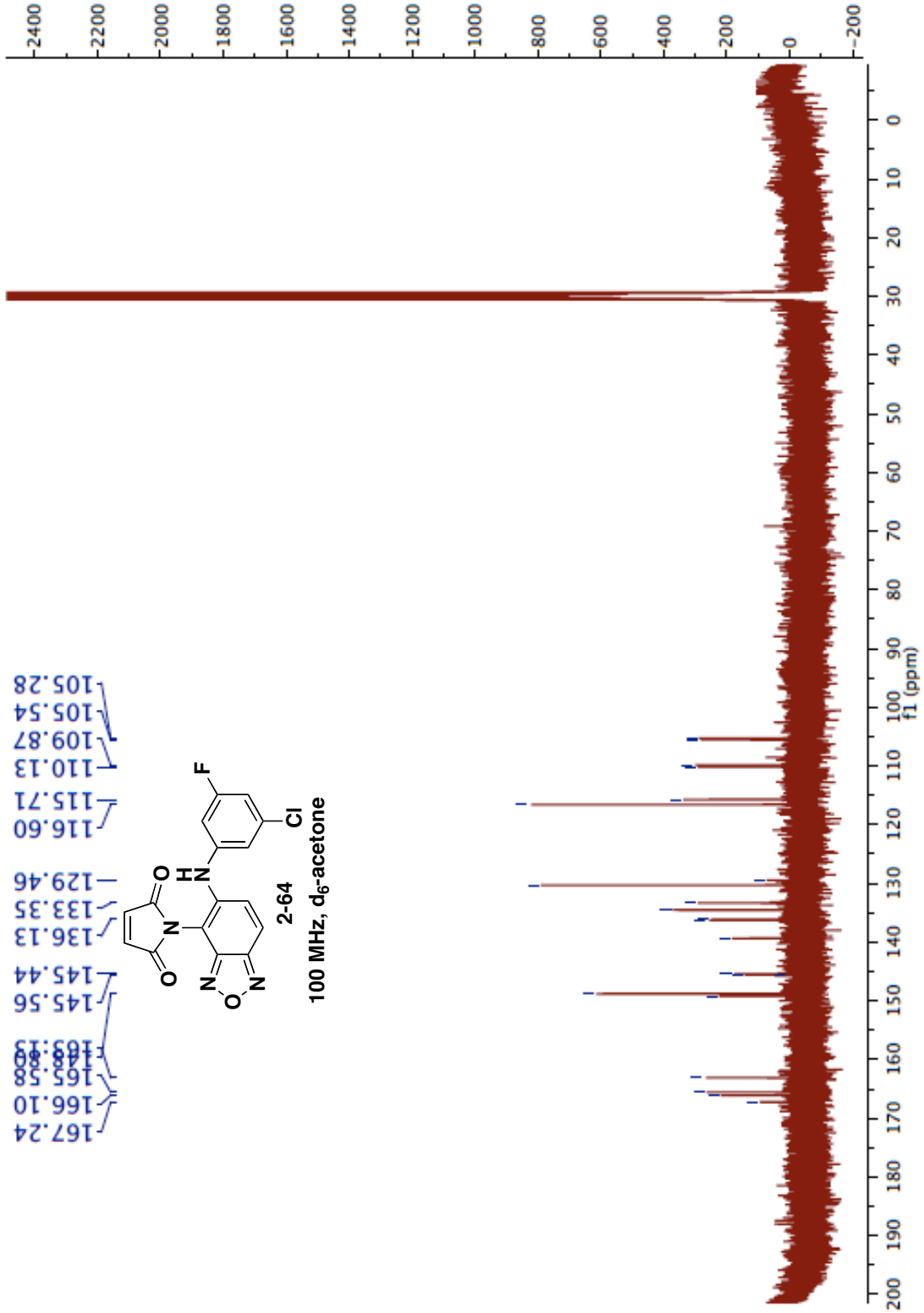

7.15
6.55
6.45
6.25
5.22
4.65
2
1
1
1
1
2
f1 (ppm)
2-57
400 MHz, CDCl3

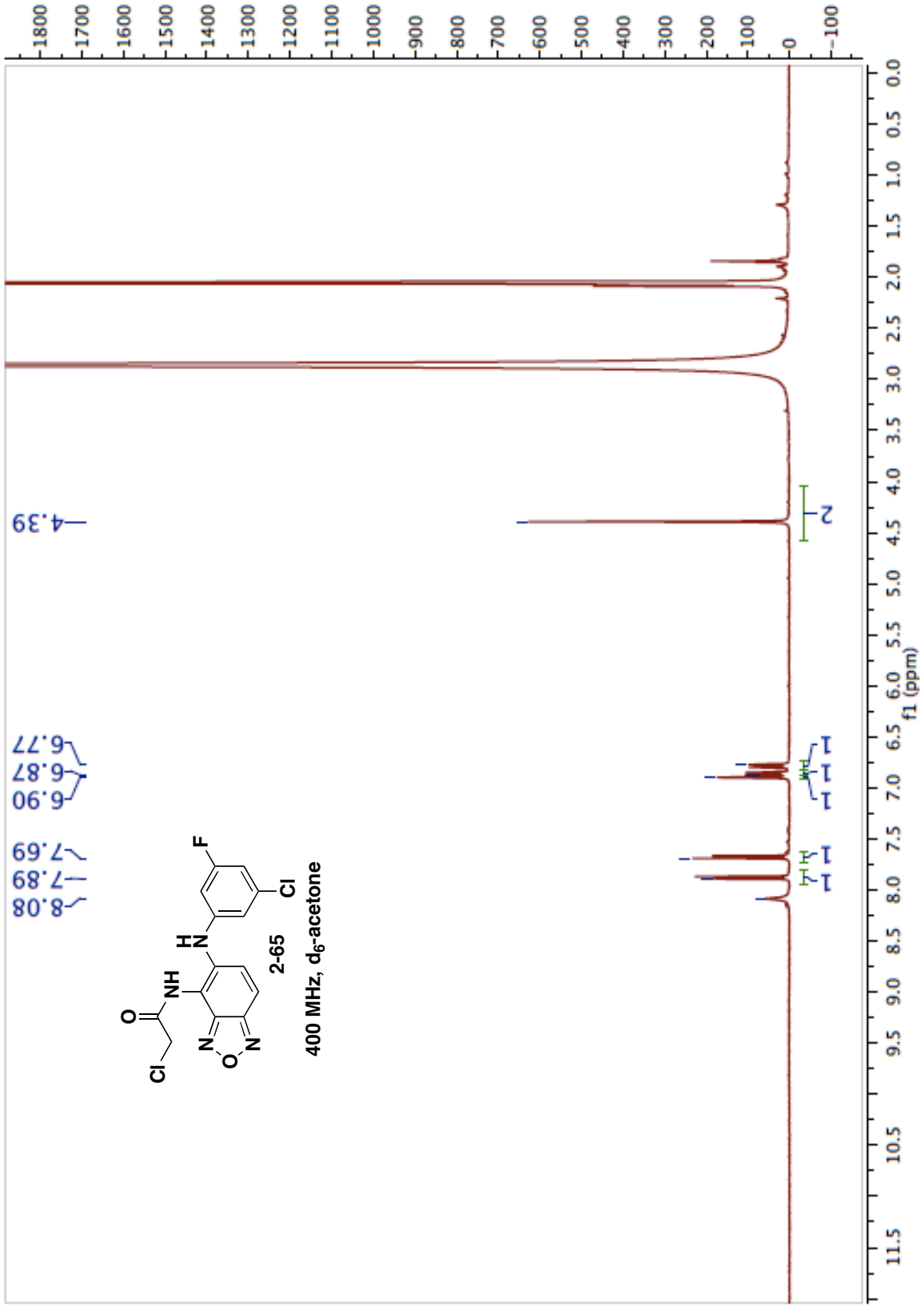

77.54
77.29
77.03
100.20
100.42
105.04
107.64
118.38
130.26
134.22
136.26
145.45
147.04
147.13
163.04
165.00
NH2
H
N
F
Cl
N
O
N
2-57
100 MHz, CDCl3
f1 (ppm)

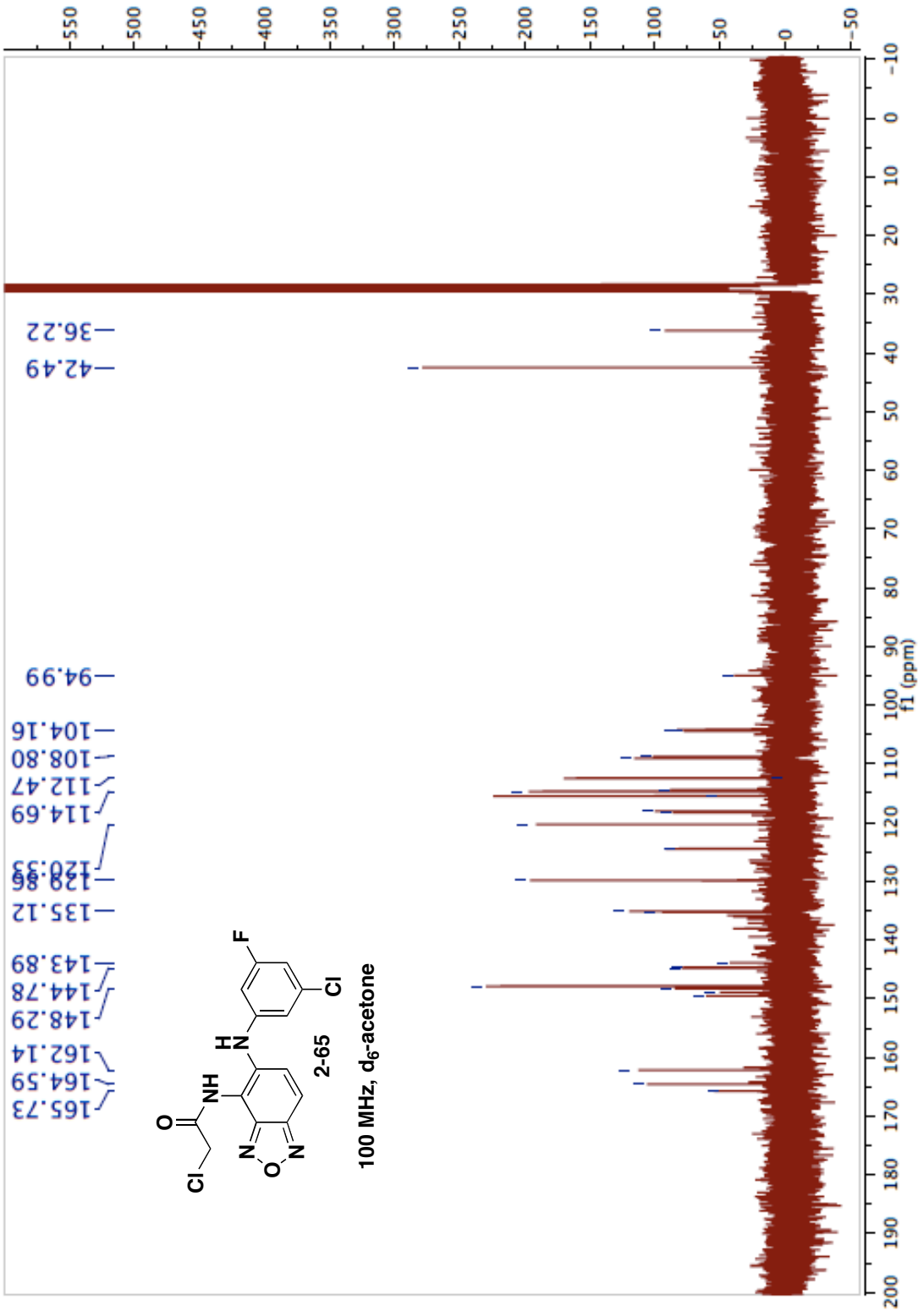

2-63
400 MHz, d6-acetone
f1 (ppm)

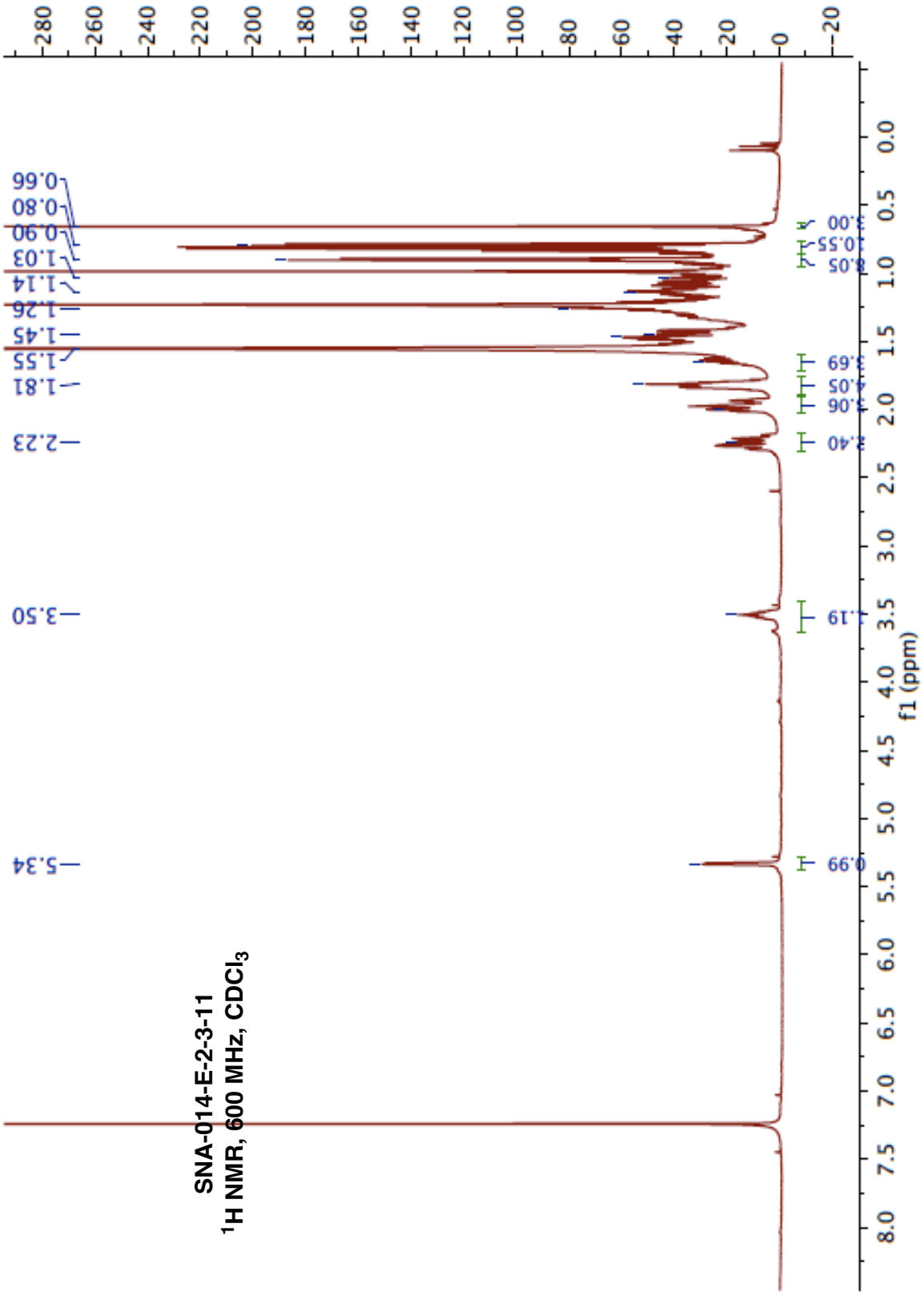

22.29
76.18
89.97
103.29
103.58
108.17
113.88
114.63
130.25
135.16
135.29
145.02
145.14
147.93
162.21
169.55
2-63
100 MHz, d6-acetone
f1 (ppm)

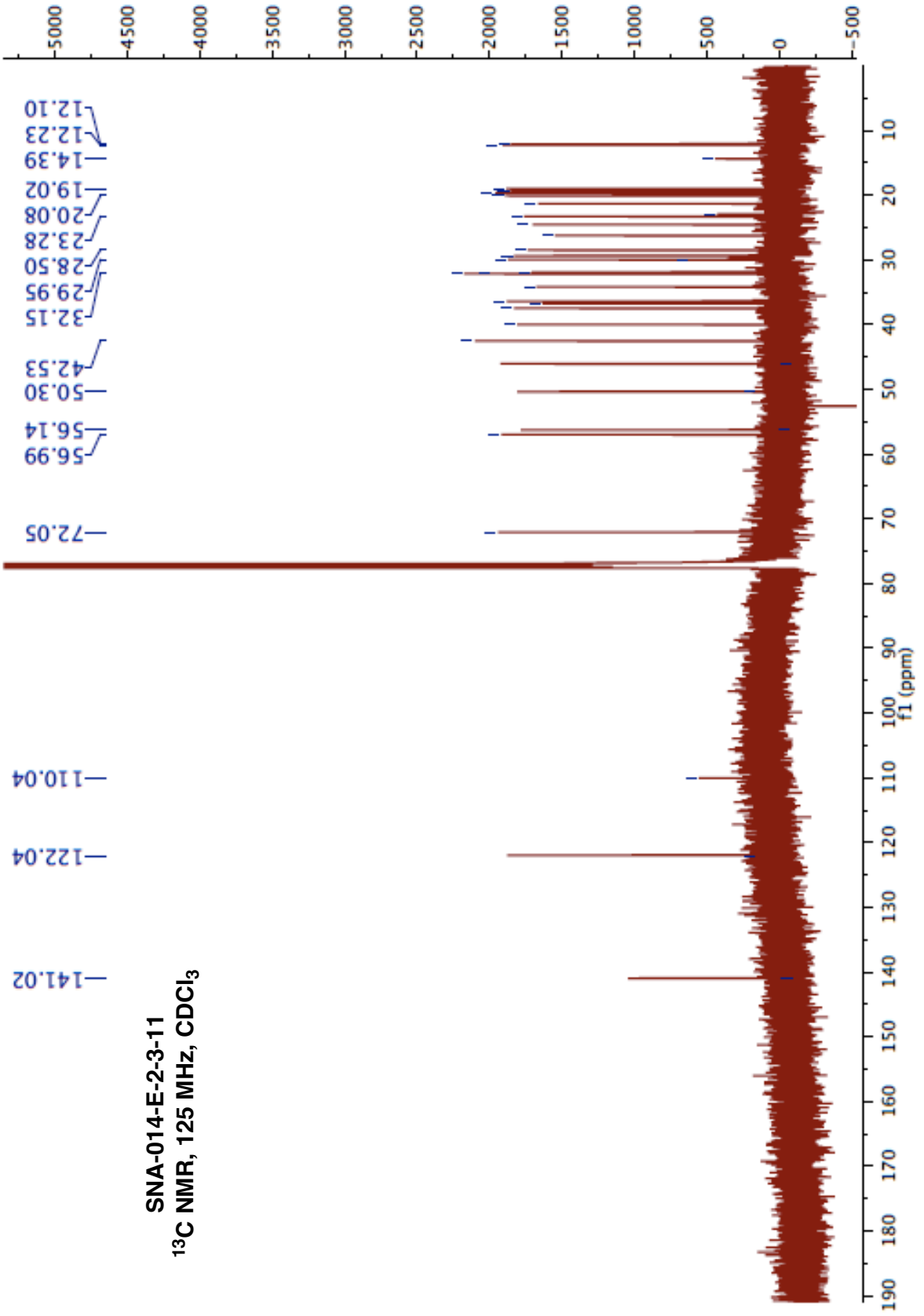

8.42
7.71
7.71
6.93
6.78
6.76
6.41
6.39
1H
3H
1H
1H
2-64
400 MHz, d6-acetone
f1 (ppm)

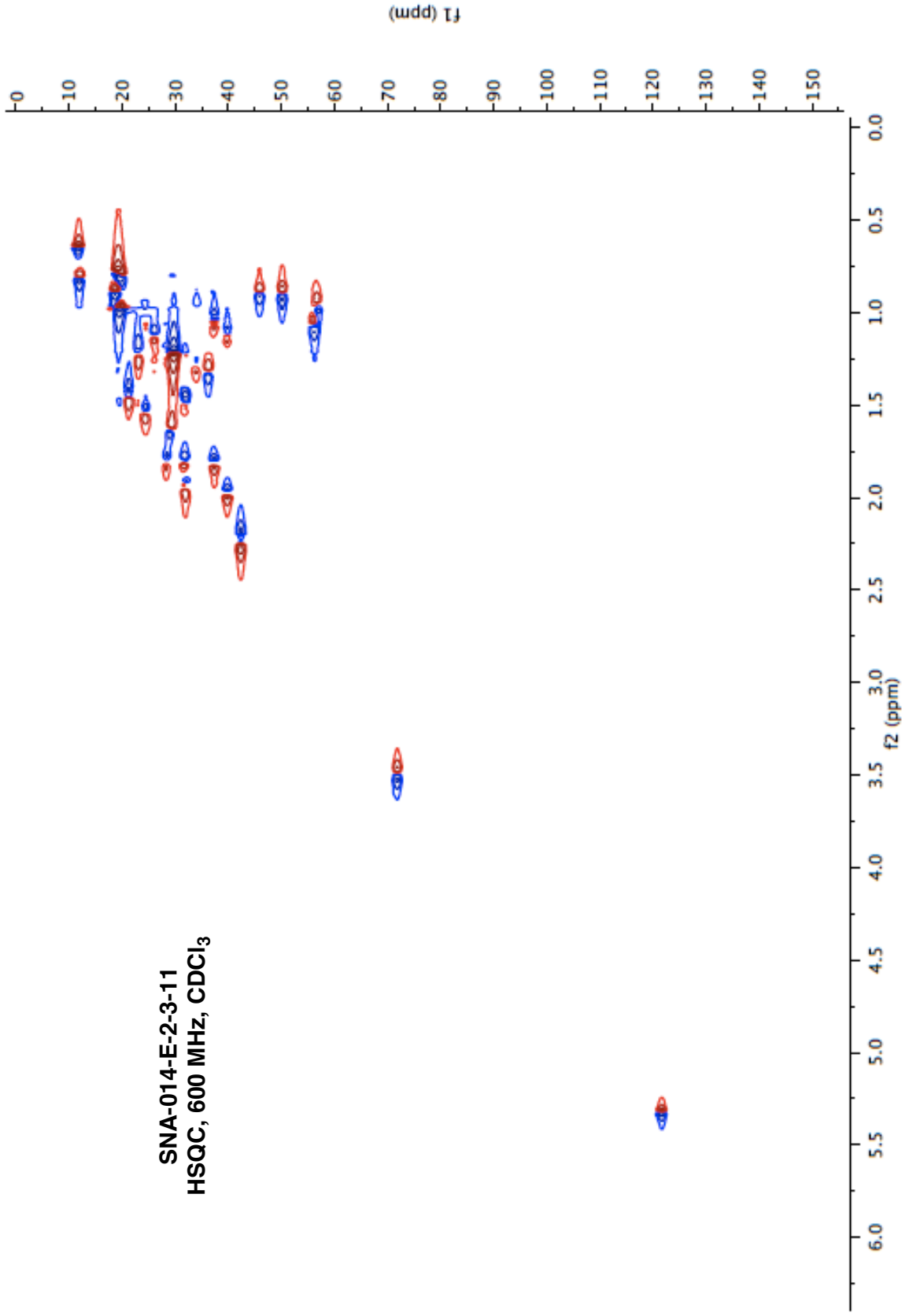

2-64
100 MHz, d6-acetone
f1 (ppm)

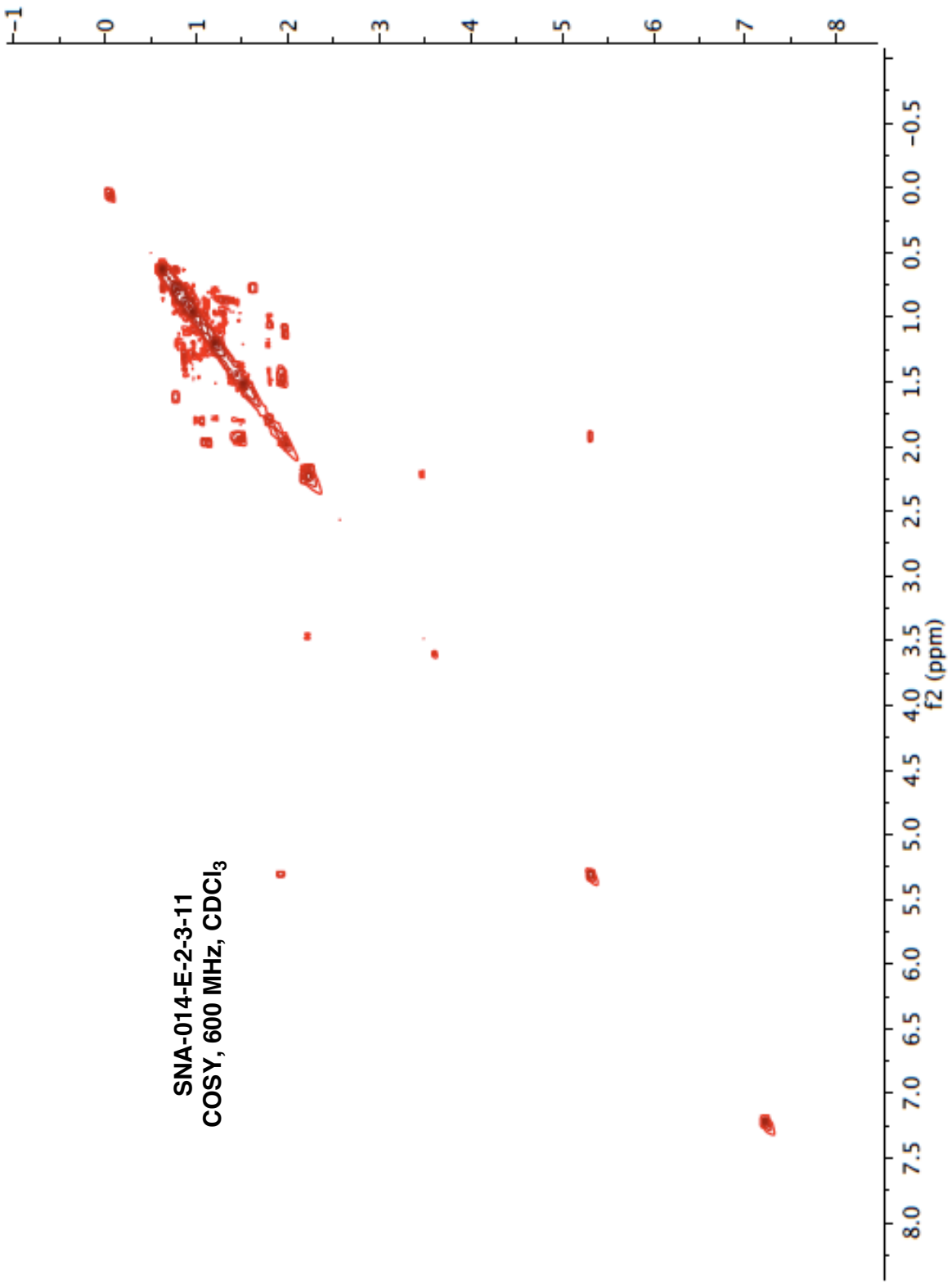

8.08
7.89
7.69
6.90
6.87
6.77
4.39
2
1
1
1
1
1
2-65
400 MHz, d6-acetone
f1 (ppm)

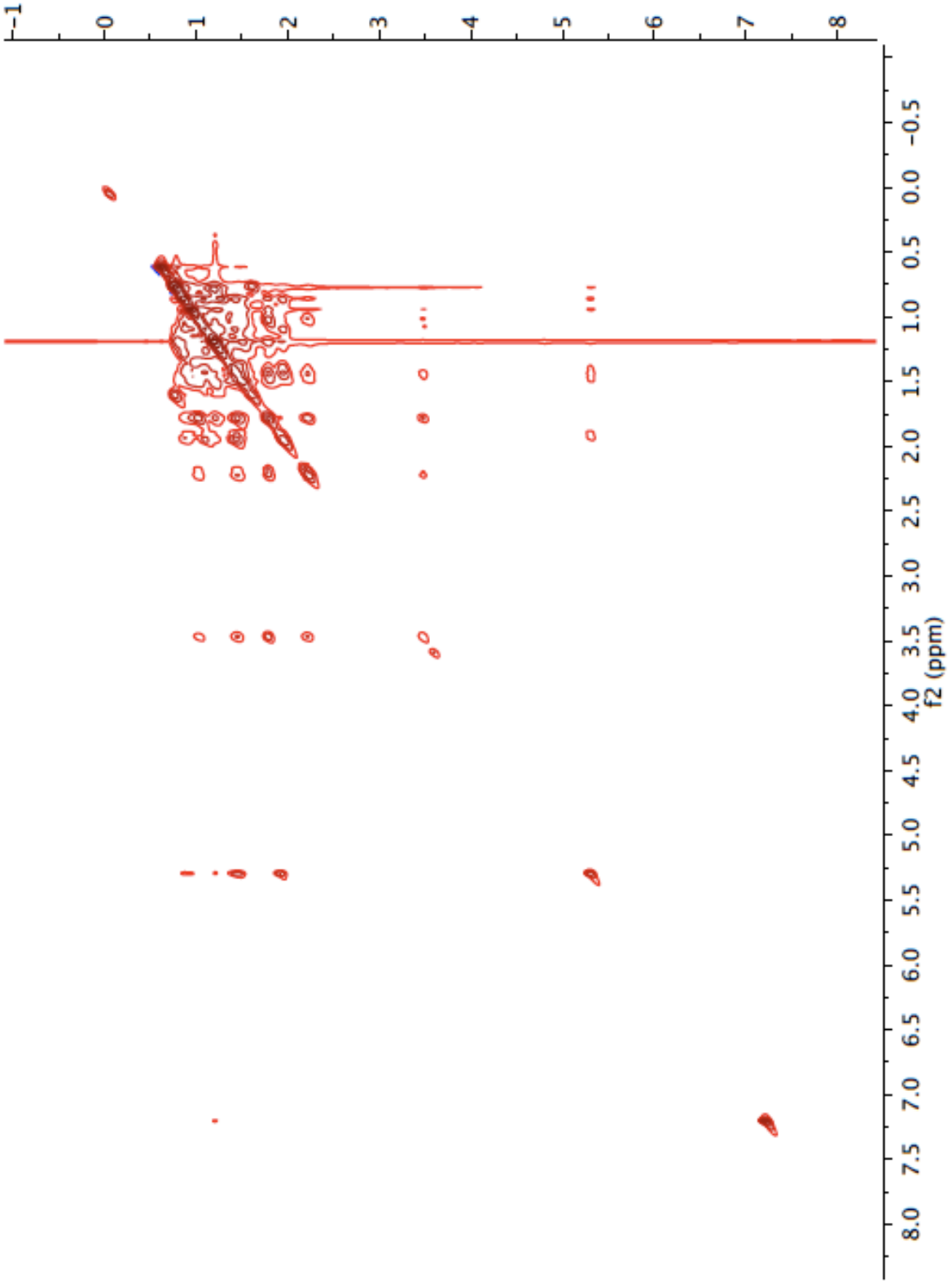

36.22
42.49
94.99
104.16
108.80
112.47
114.69
120.39
129.86
135.12
143.89
144.78
148.29
162.14
164.59
165.73
F
Cl
H
N
NH
O
Cl
N
O
N
2-65
100 MHz, d6-acetone
f1 (ppm)

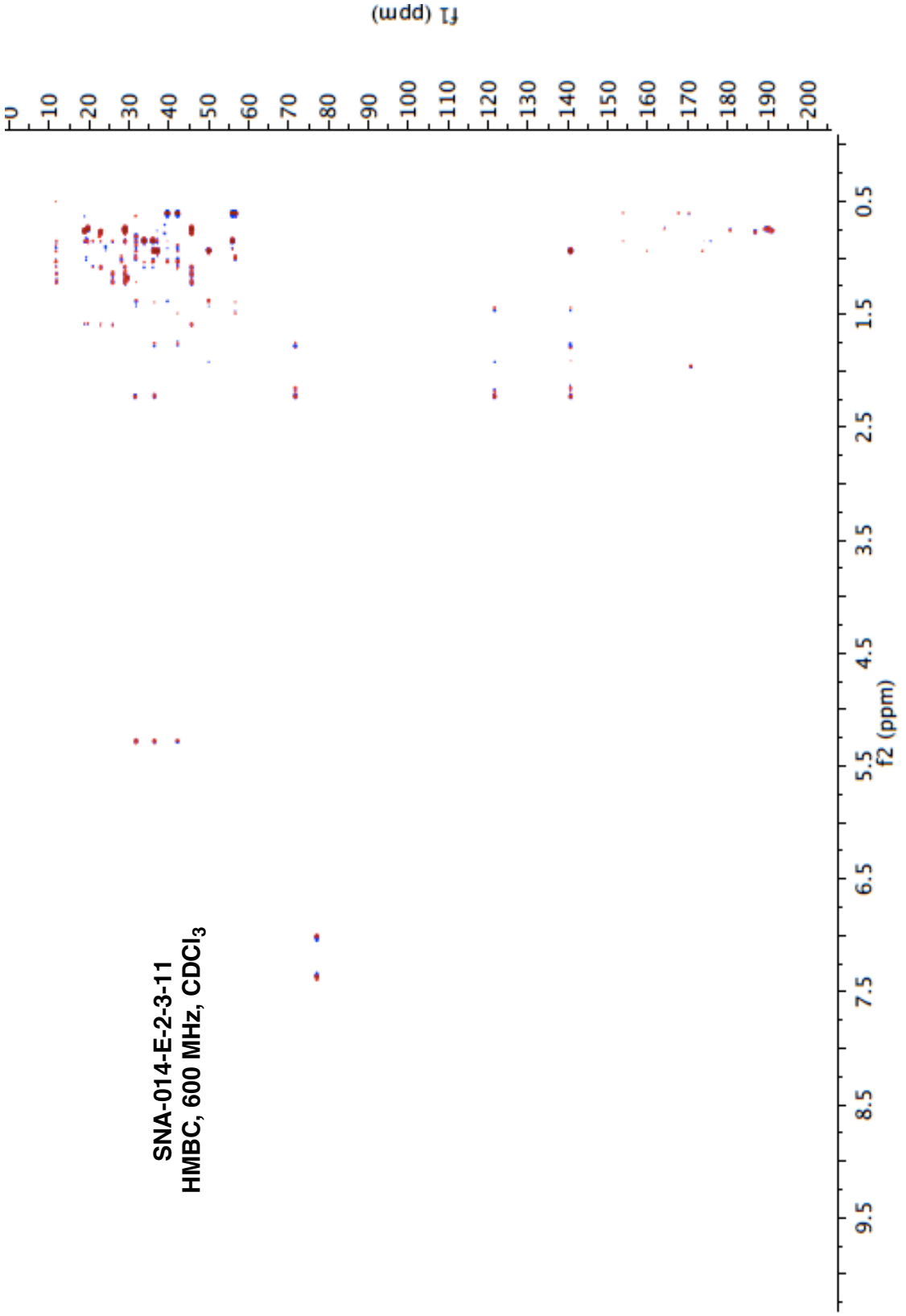

SNA-014-E-2-3-11
^{1}H NMR, 600 MHz, $CDCl_3$
f1 (ppm)
5.34
3.50
2.23
1.81
1.55
1.45
1.26
1.14
1.03
0.90
0.80
0.66
0.99
1.19
2.40
3.06
4.05
3.69
8.05
10.55
3.00

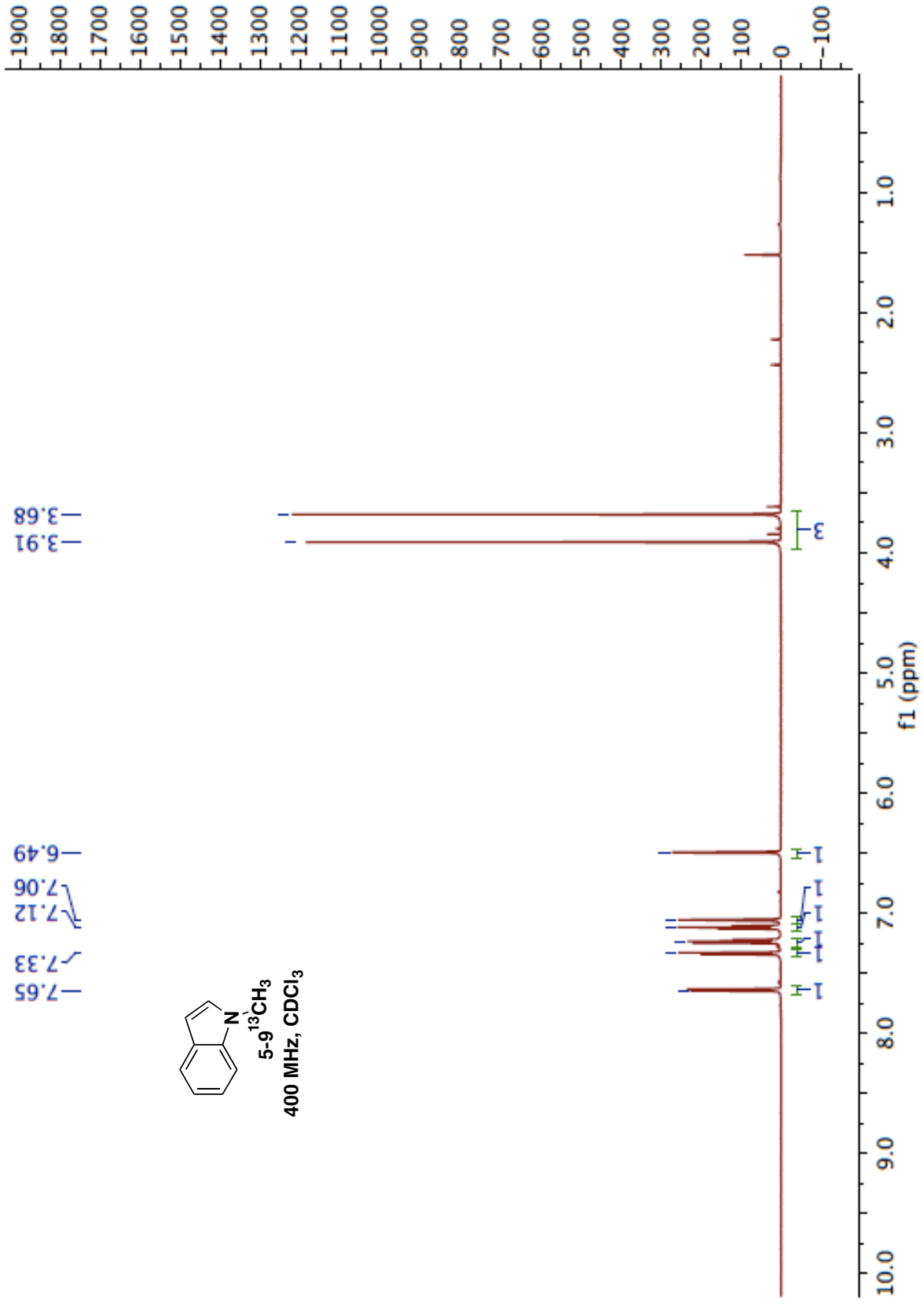

SNA-014-E-2-3-11
^{13}C NMR, 125 MHz, $CDCl_3$
141.02
122.04
110.04
72.05
56.99
56.14
50.30
42.53
32.15
29.95
28.50
23.28
20.08
19.02
14.39
12.23
12.10
f1 (ppm)

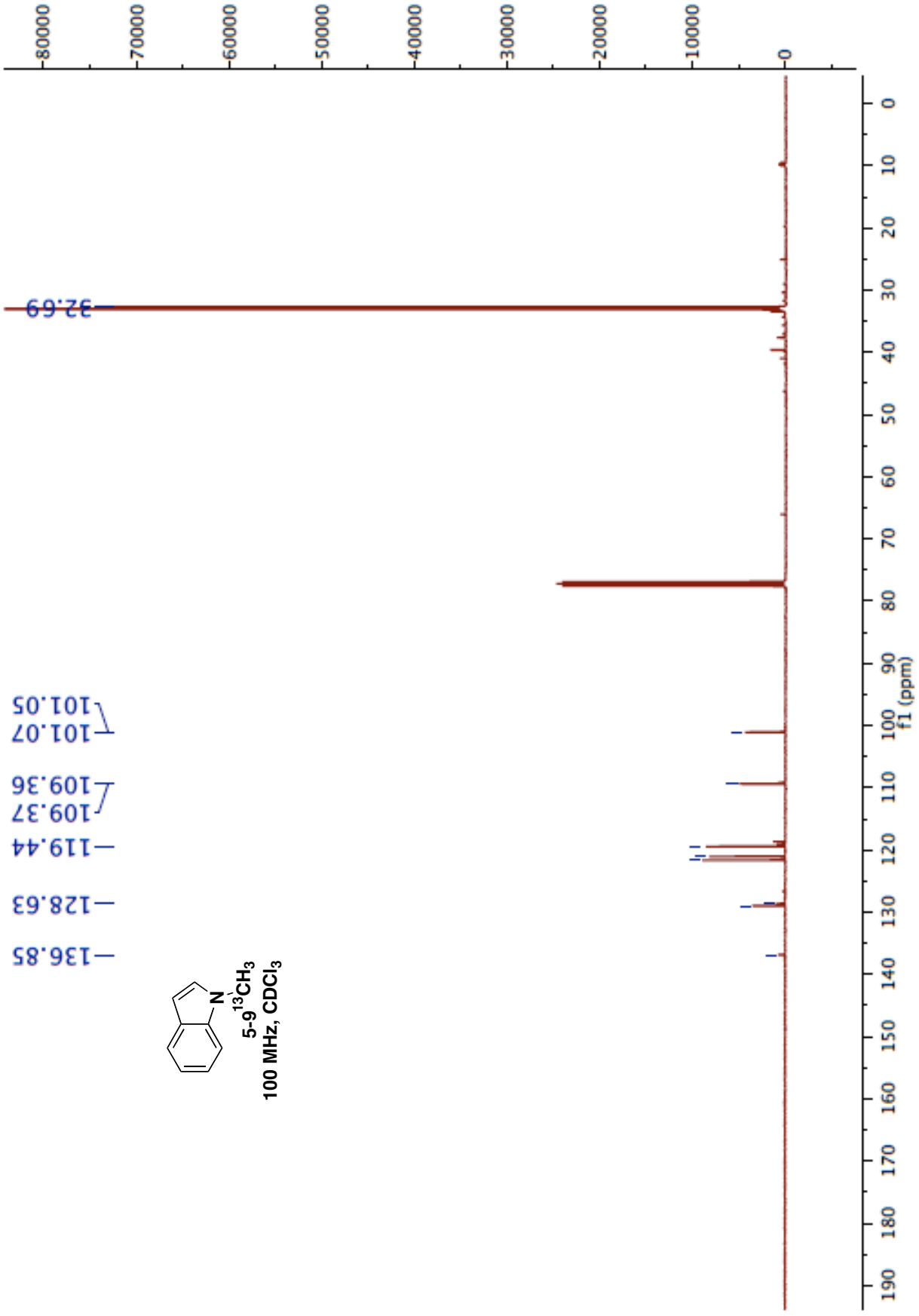

SNA-014-E-2-3-11
HSQC, 600 MHz, CDCl3
f1 (ppm)
f2 (ppm)

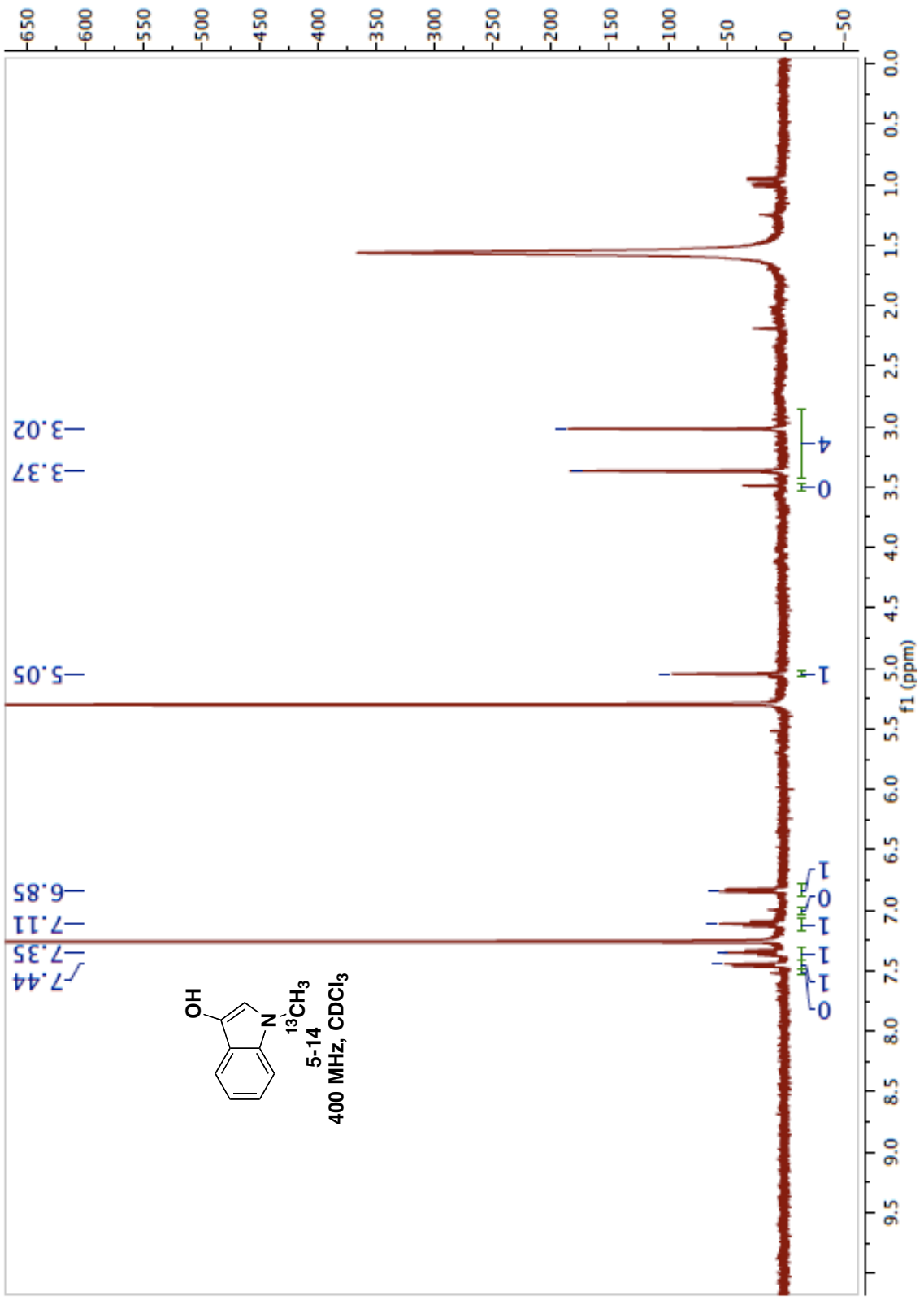

SNA-014-E-2-3-11
COSY, 600 MHz, CDCl3
f2 (ppm)
f1 (ppm)

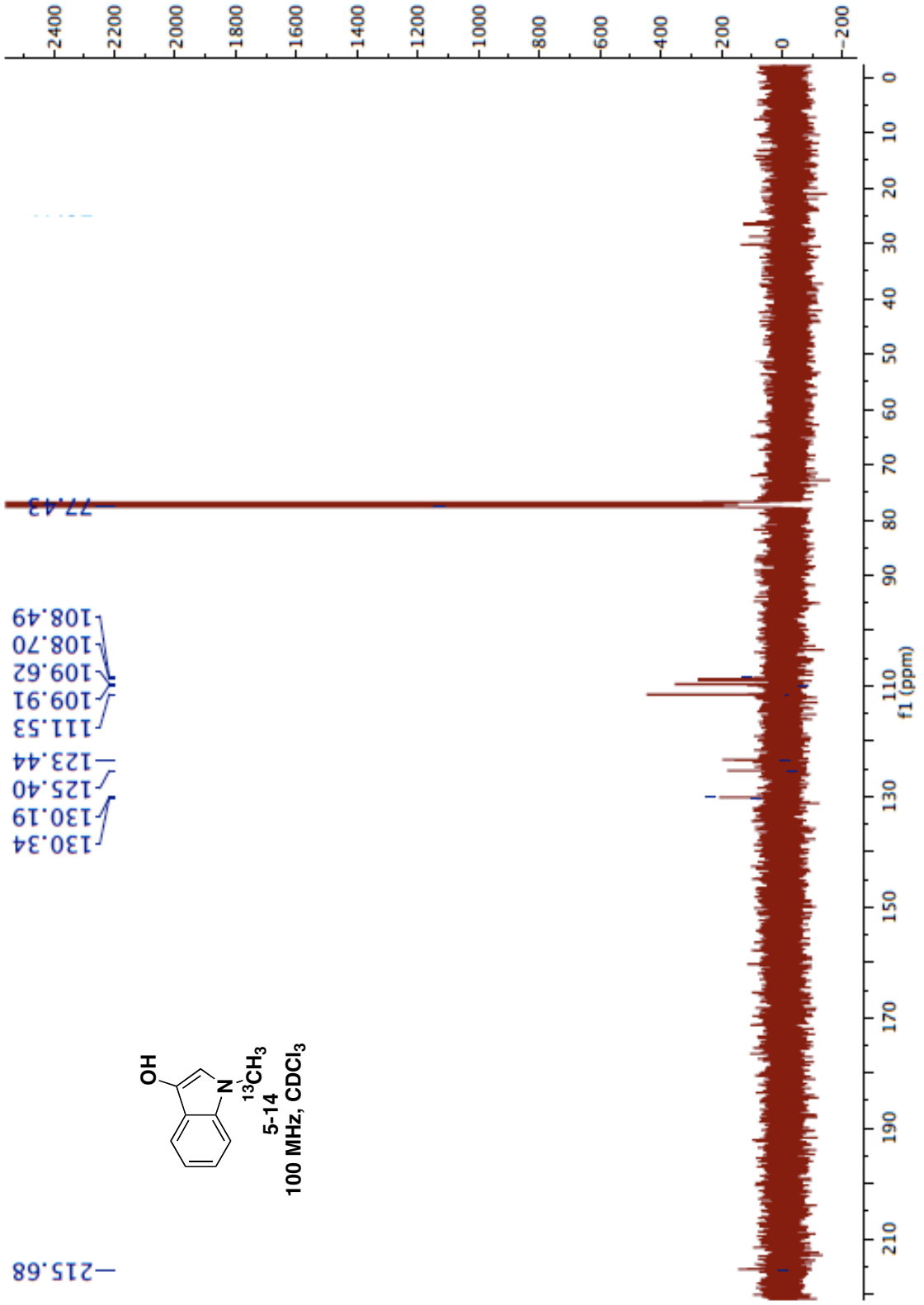

f1 (ppm)
f2 (ppm)

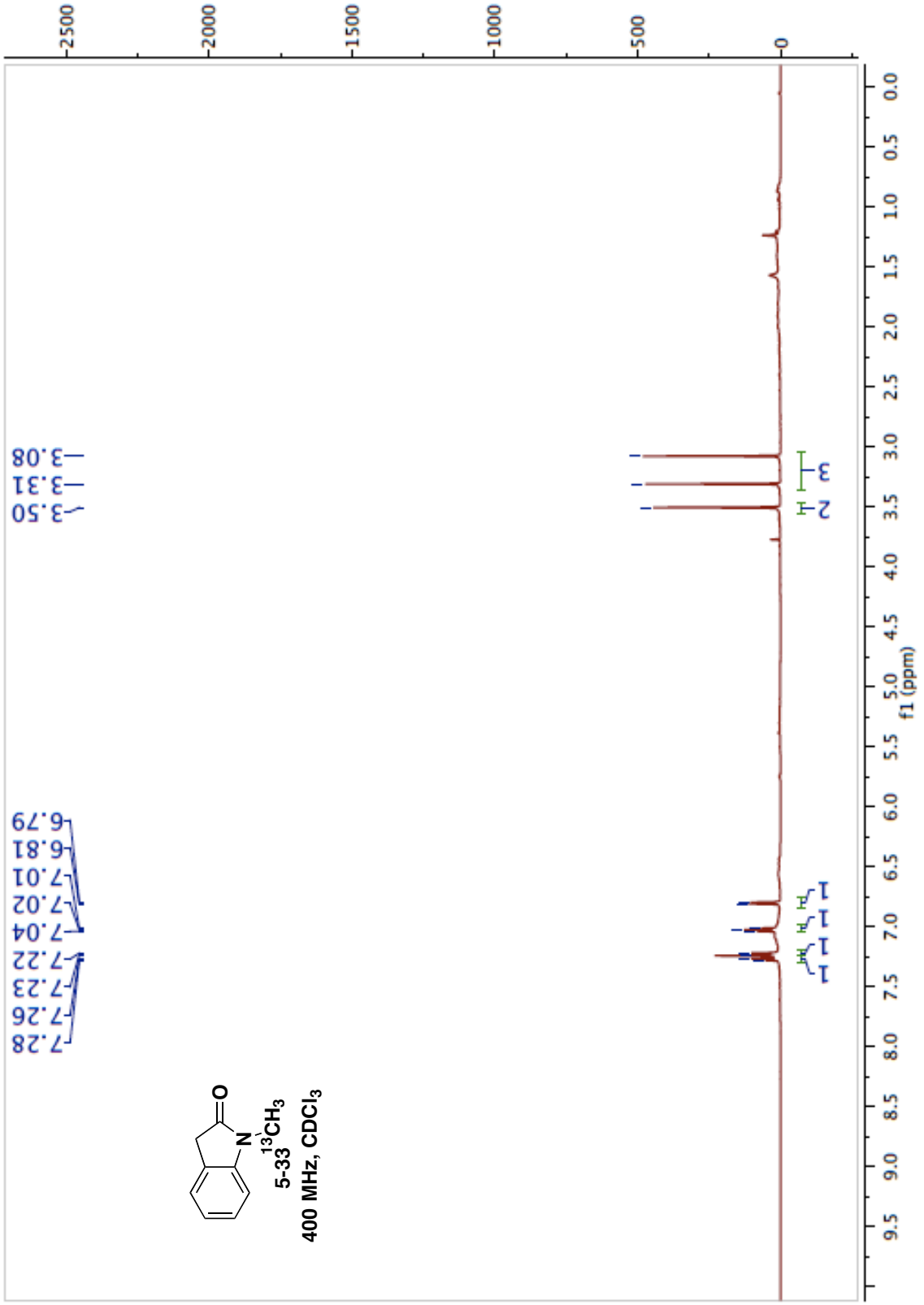

SNA-014-E-2-3-11
HMBC, 600 MHz, CDCl3
f1 (ppm)
f2 (ppm)

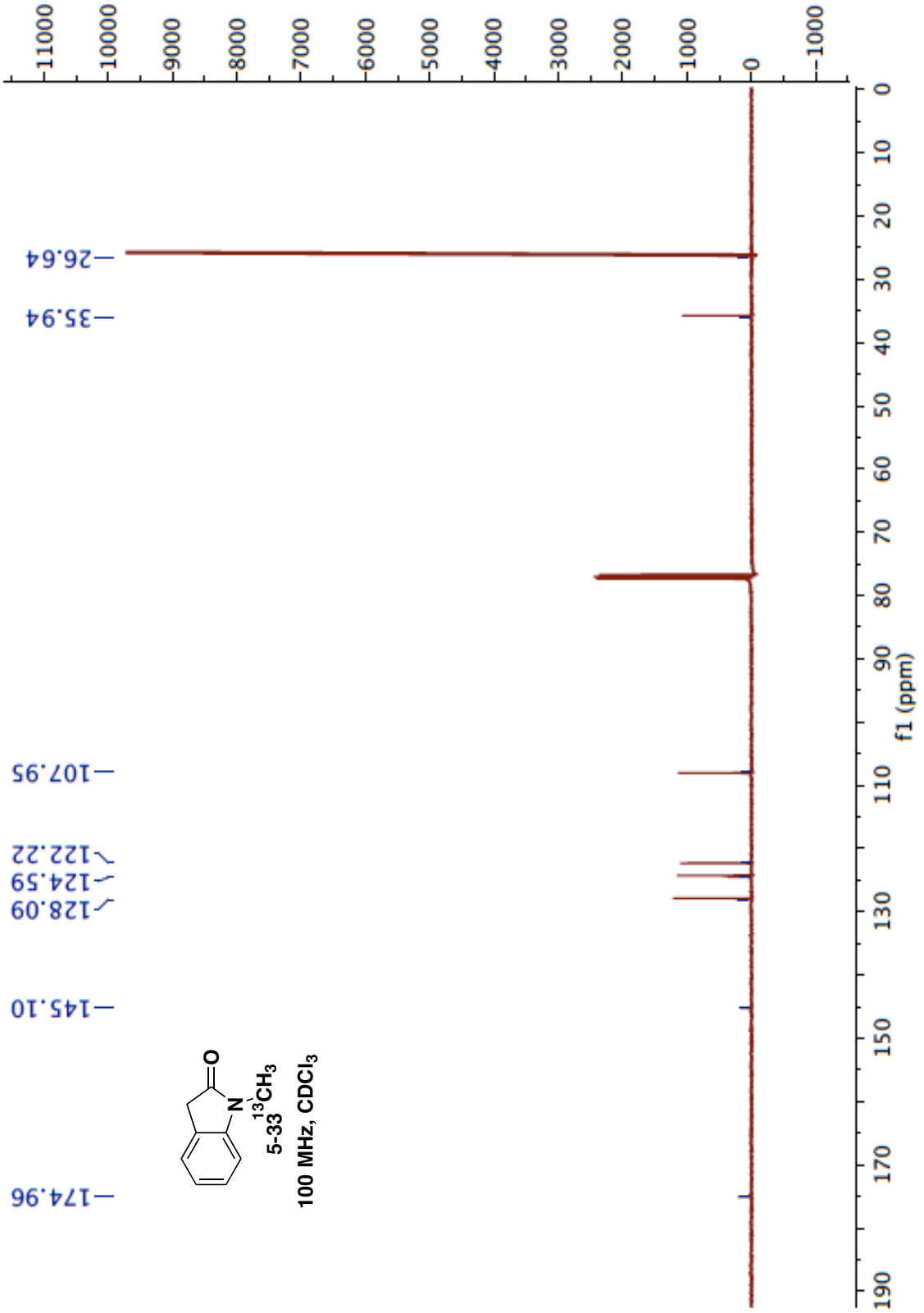

5-9 $^{13}CH_3$
400 MHz, $CDCl_3$
3.68
3.91
6.49
7.06
7.12
7.33
7.65
3
1
1
1
1
1
1
f1 (ppm)

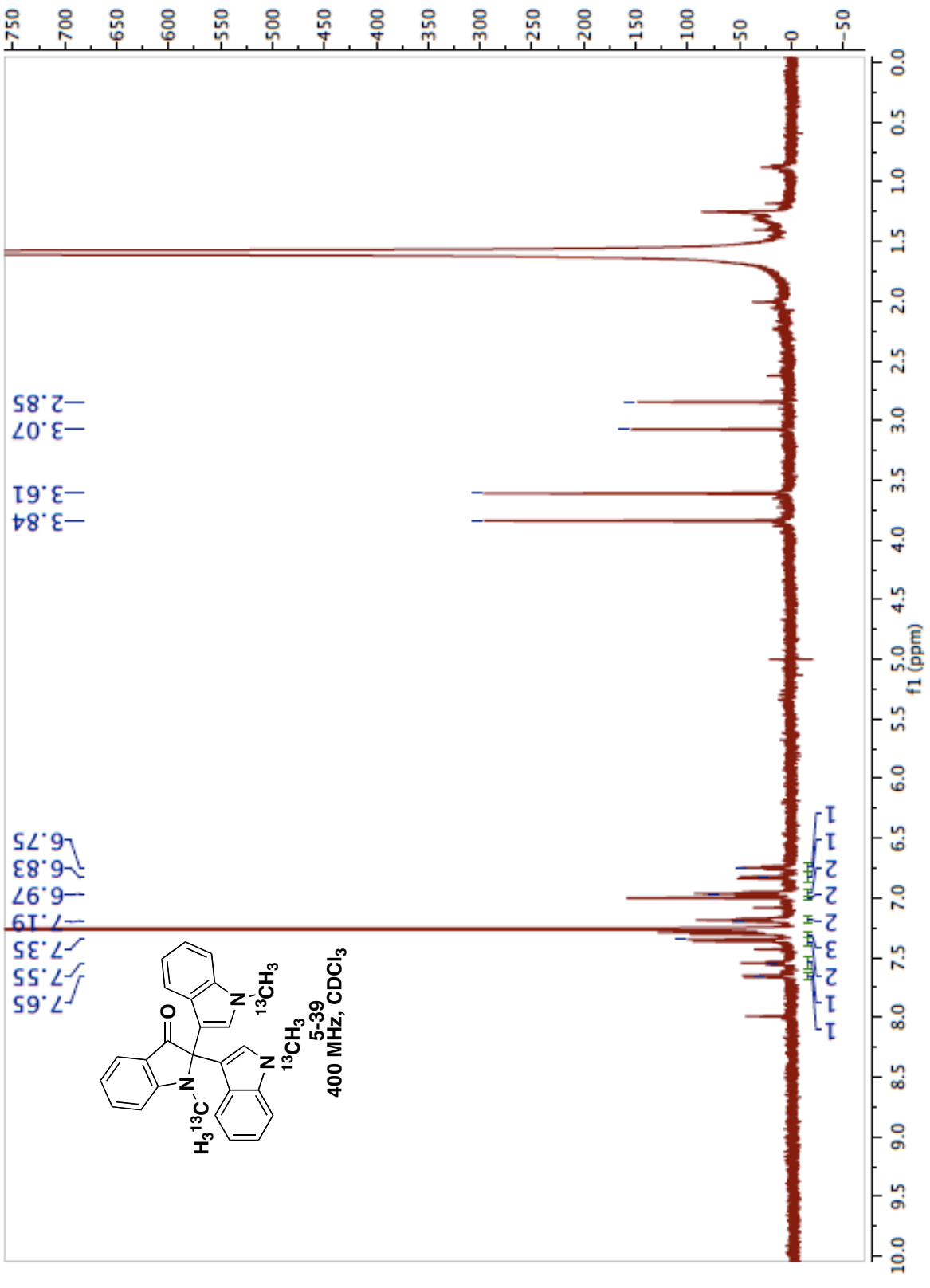

32.69
101.05
101.07
109.36
109.37
119.44
128.63
136.85
N
$^{13}CH_3$
5-9
100 MHz, $CDCl_3$
f1 (ppm)

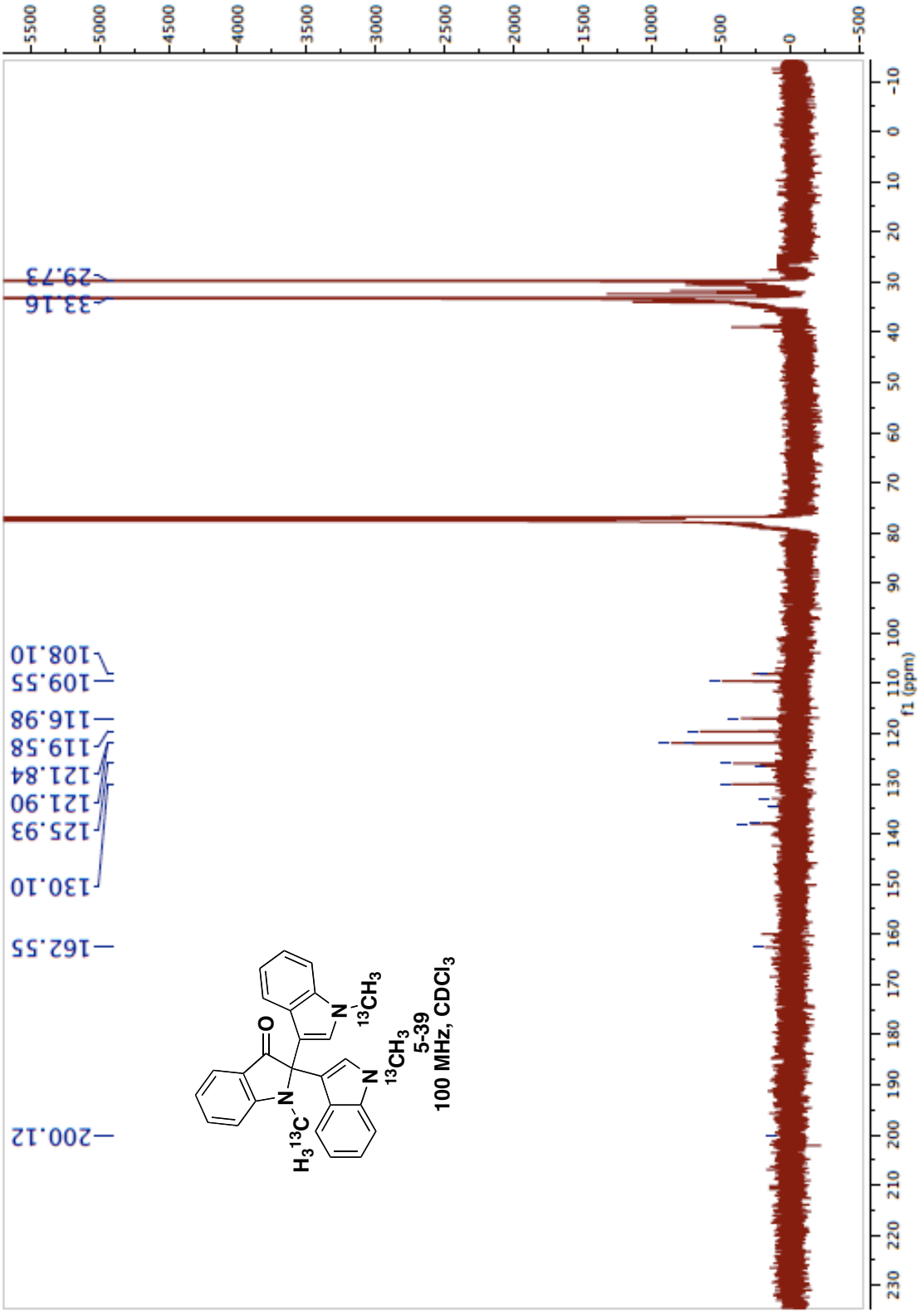

OH
N
$^{13}CH_3$
5-14
400 MHz, $CDCl_3$
7.44
7.35
7.11
6.85
5.05
3.37
3.02
f1 (ppm)

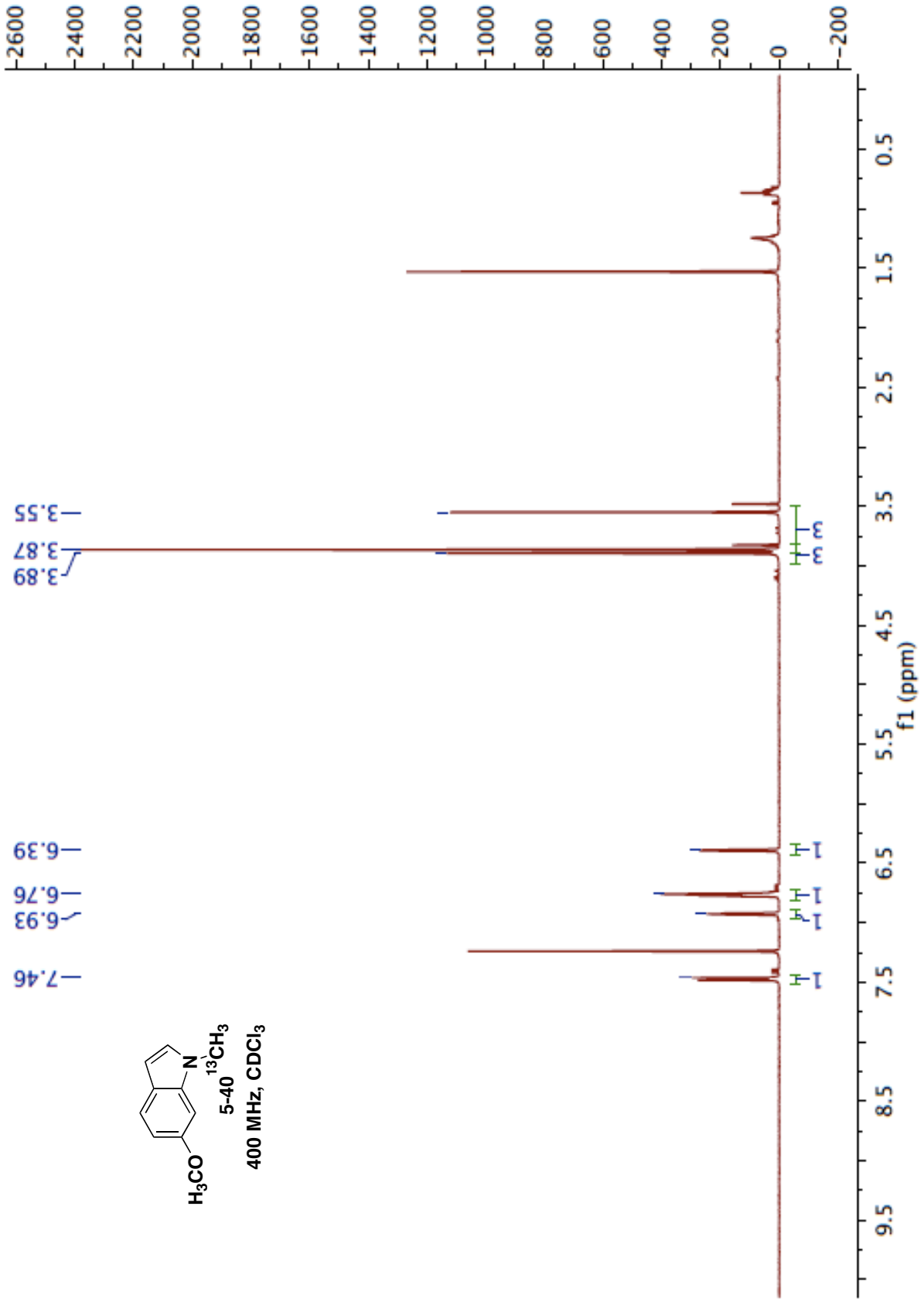

215.68
130.34
130.19
125.40
123.44
111.53
109.91
109.62
108.70
108.49
77.43
OH
N
$^{13}CH_3$
5-14
100 MHz, $CDCl_3$
f1 (ppm)

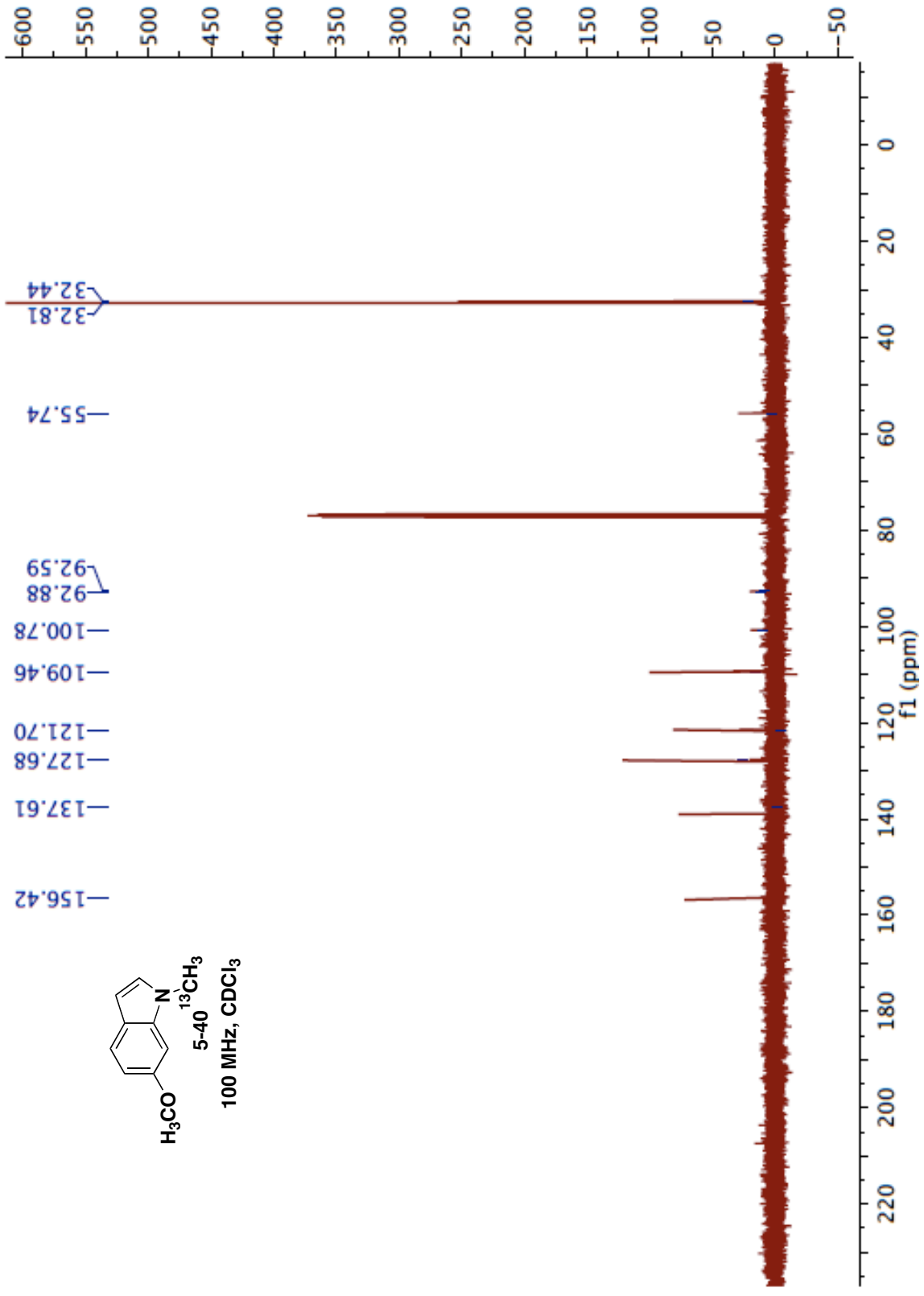

5-33
400 MHz, CDCl3
3.08
3.31
3.50
6.79
6.81
7.01
7.02
7.04
7.22
7.23
7.26
7.28
f1 (ppm)

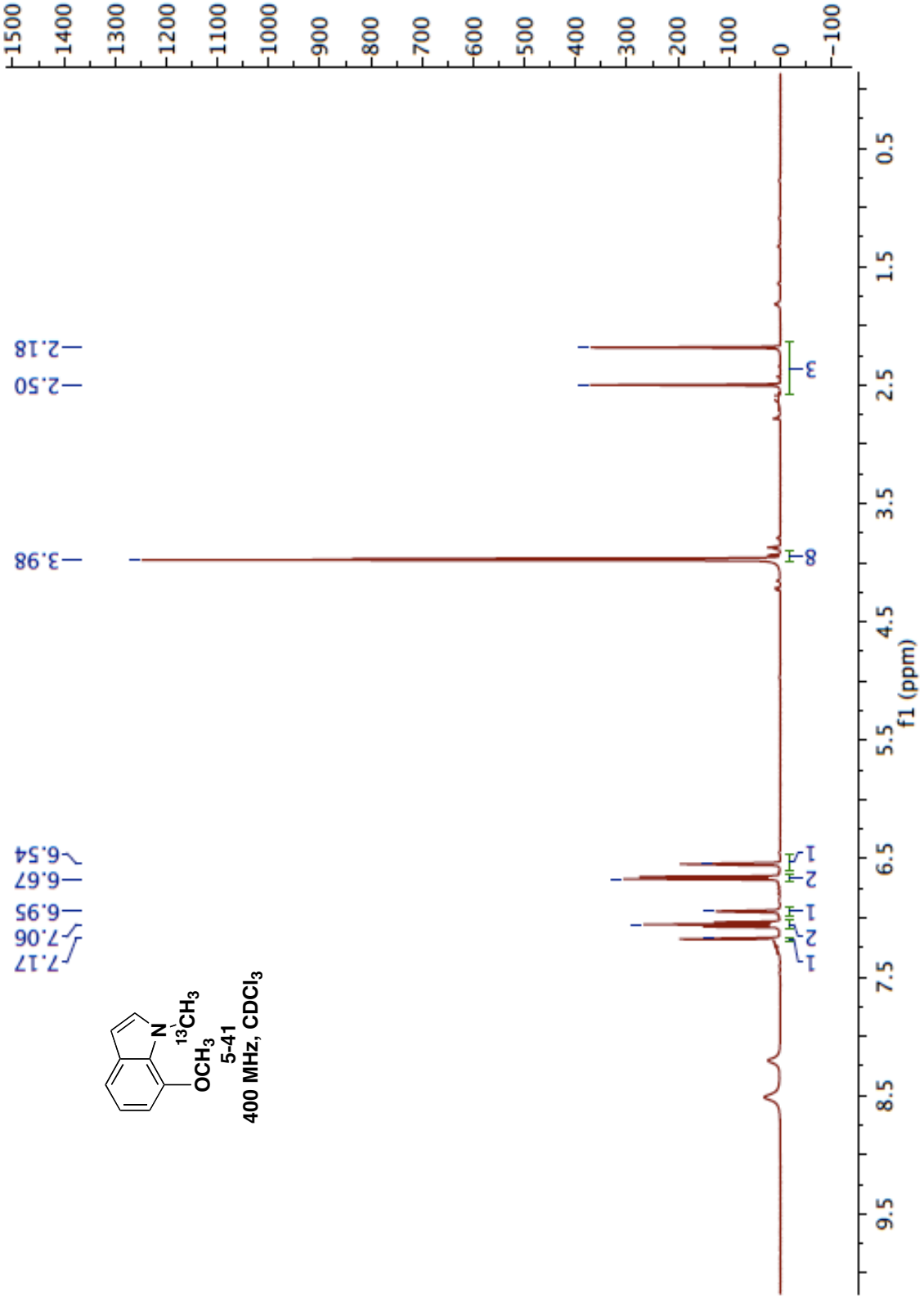

174.96
145.10
128.09
124.59
122.22
107.95
35.94
26.64
5-33
100 MHz, CDCl3
f1 (ppm)

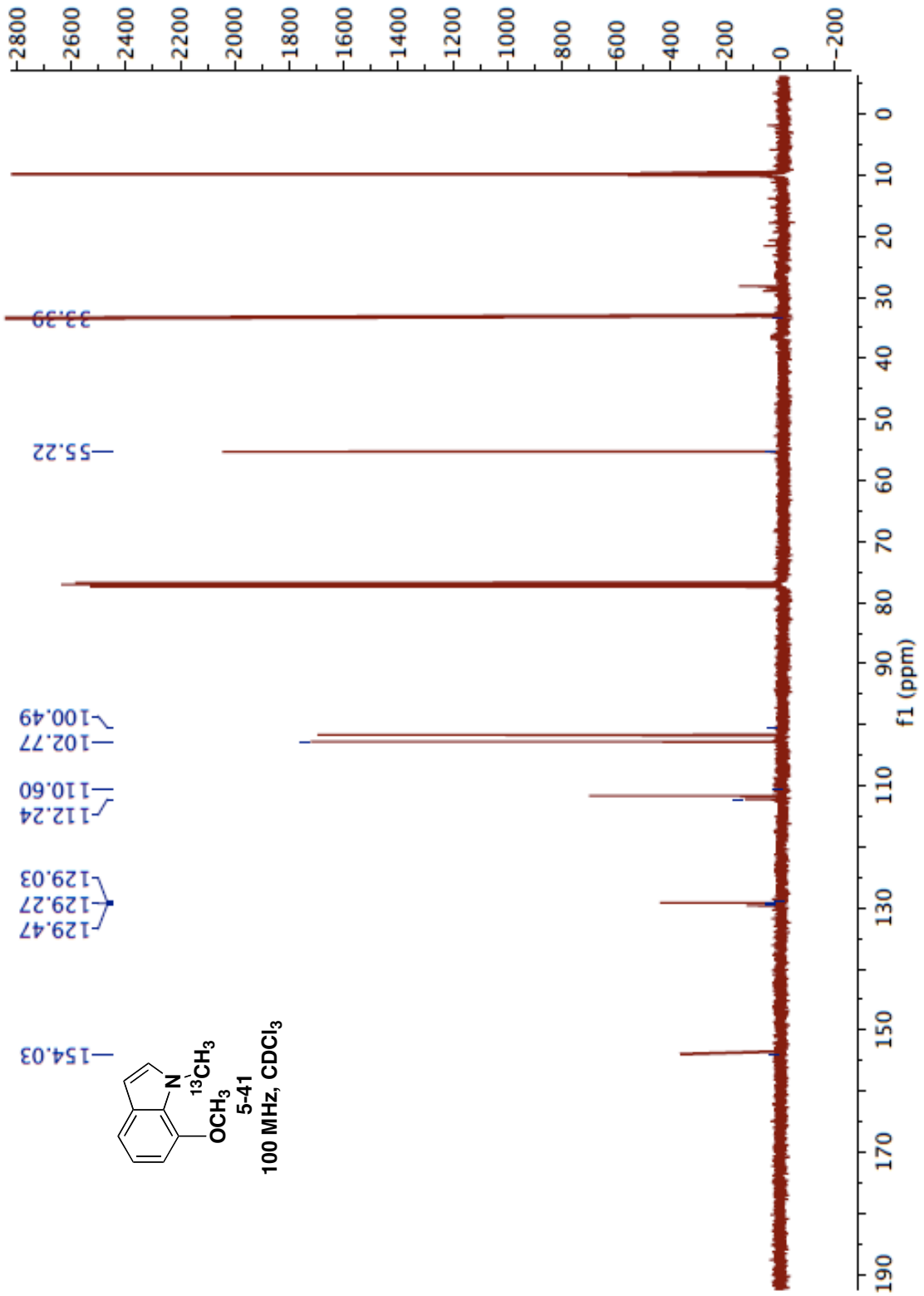

7.65
7.55
7.35
7.19
6.97
6.83
6.75
3.84
3.61
3.07
2.85
5-39
400 MHz, CDCl3
f1 (ppm)

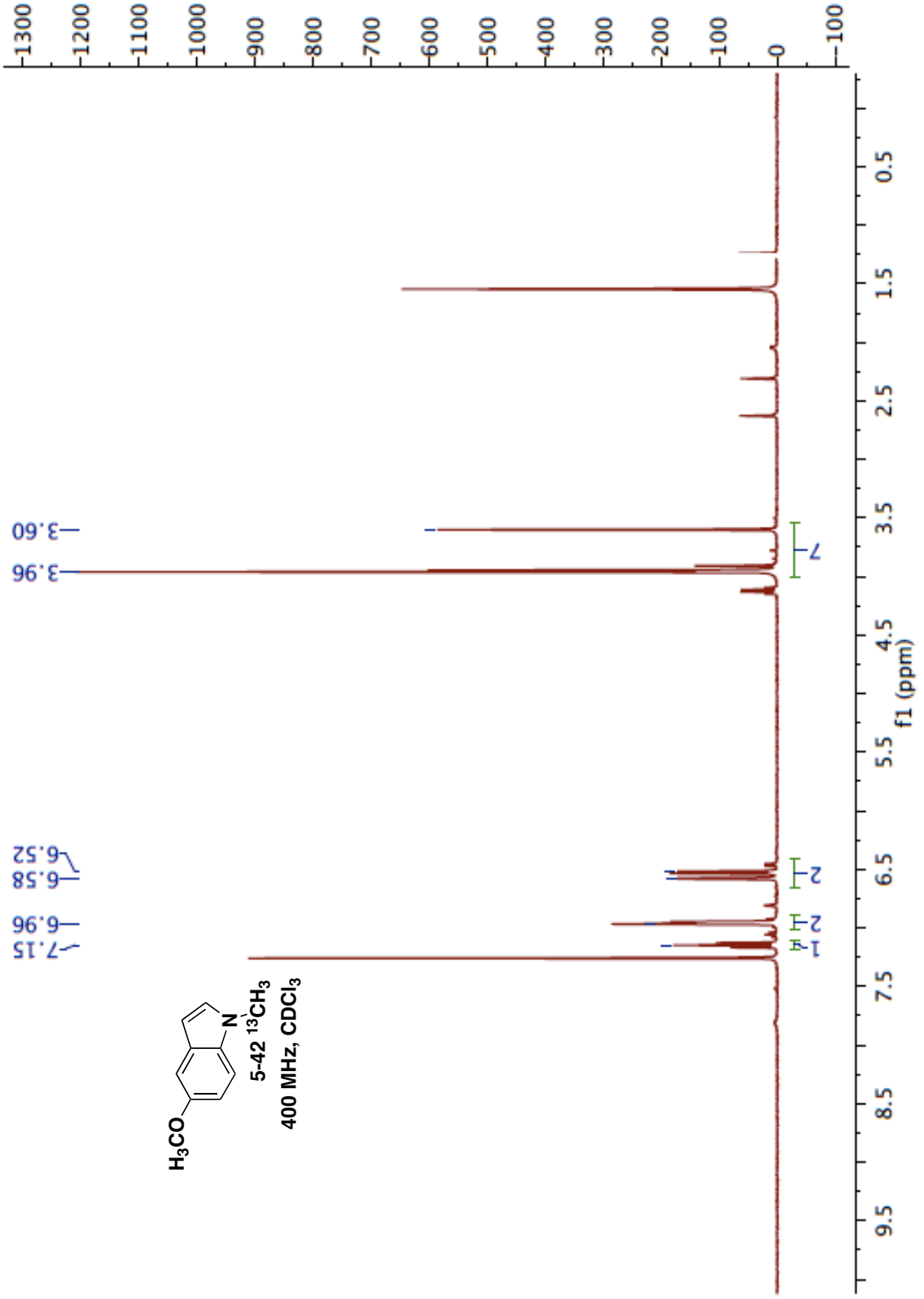

29.73
33.16
108.10
109.55
116.98
119.58
121.84
121.90
125.93
130.10
162.55
200.12
O
N
$^{13}CH_3$
$H_3{}^{13}C$
5-39
100 MHz, $CDCl_3$
f1 (ppm)

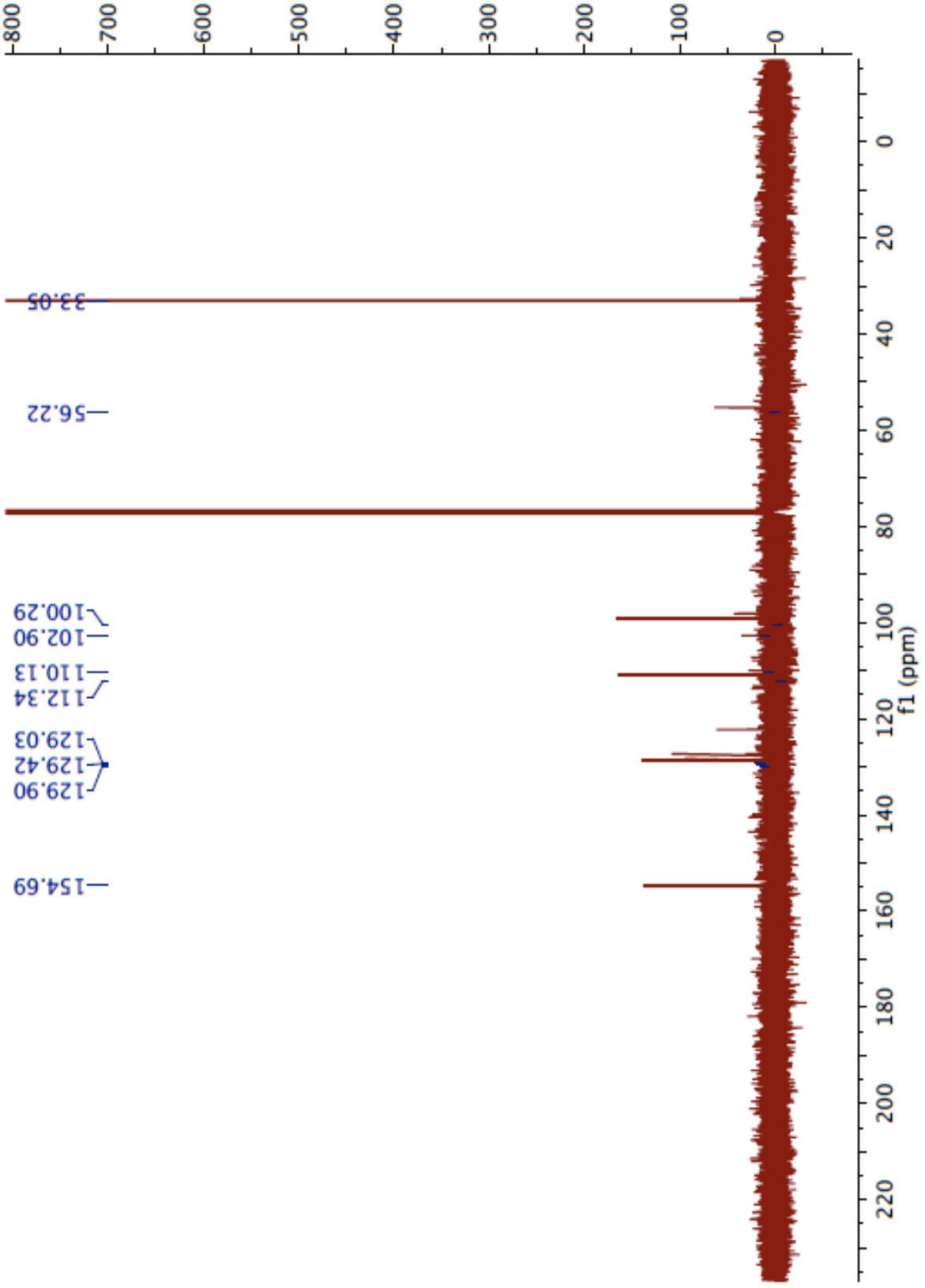

3.55
3.87
3.89
6.39
6.76
6.93
7.46
3
3
1
1
1
1
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
−200
0.5
1.5
2.5
3.5
4.5
5.5
6.5
7.5
8.5
9.5
f1 (ppm)
H3CO
N
13CH3
5-40
400 MHz, CDCl3

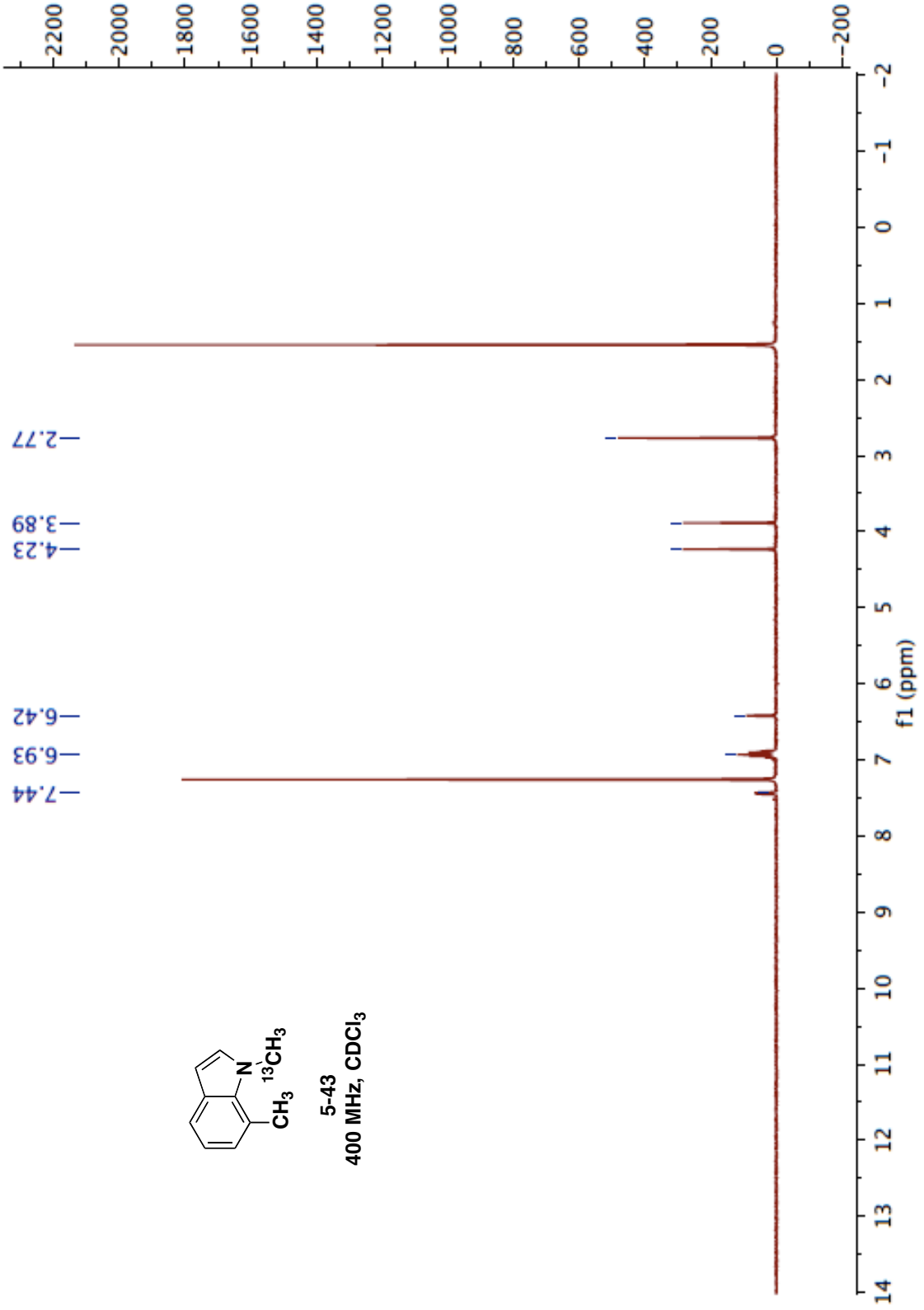

32.44
32.81
55.74
92.59
92.88
100.78
109.46
121.70
127.68
137.61
156.42
f1 (ppm)
5-40
100 MHz, CDCl3

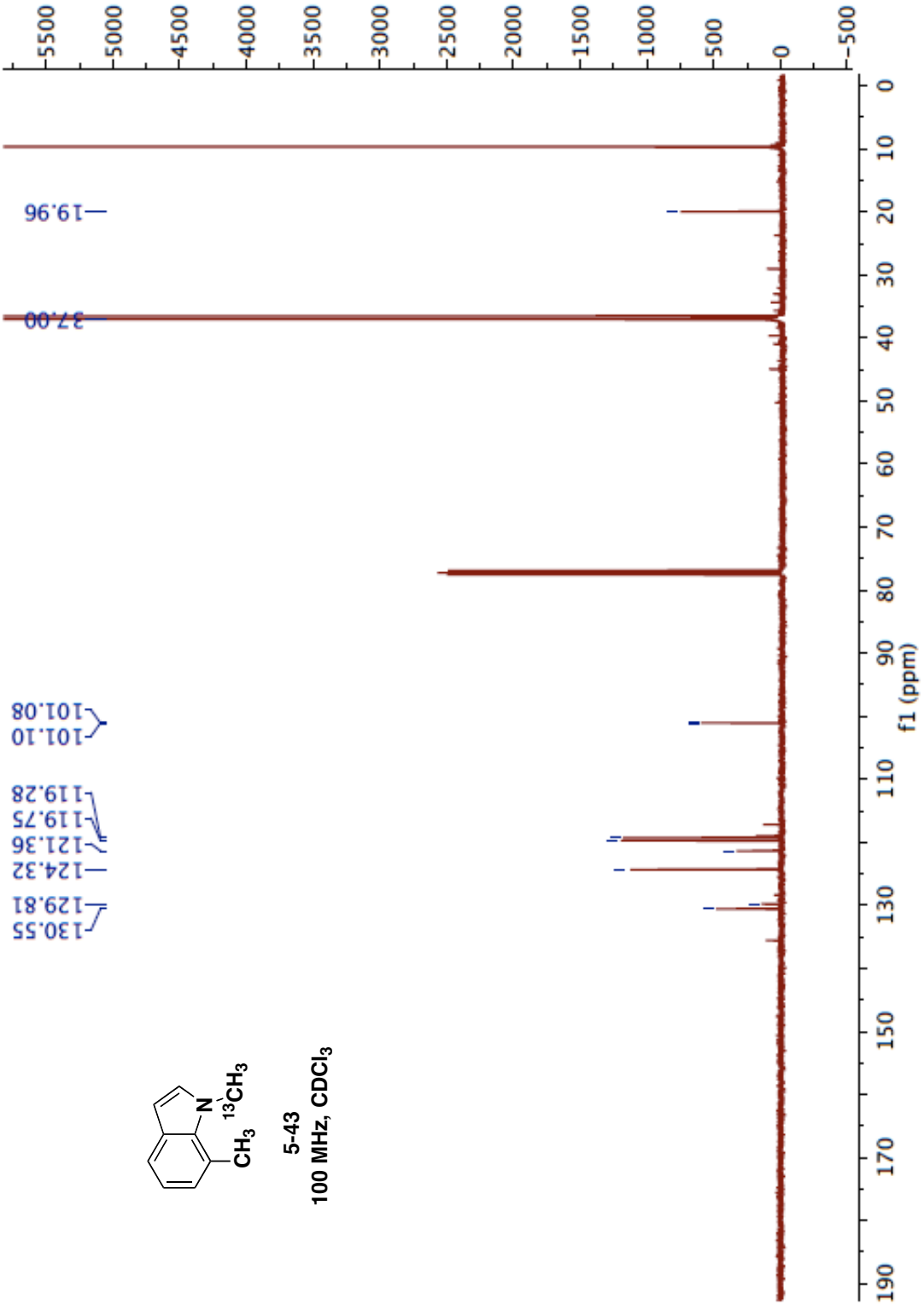

7.17
7.06
6.95
6.67
6.54
3.98
2.50
2.18
1
2
1
2
1
3
8
f1 (ppm)
5-41
400 MHz, CDCl3

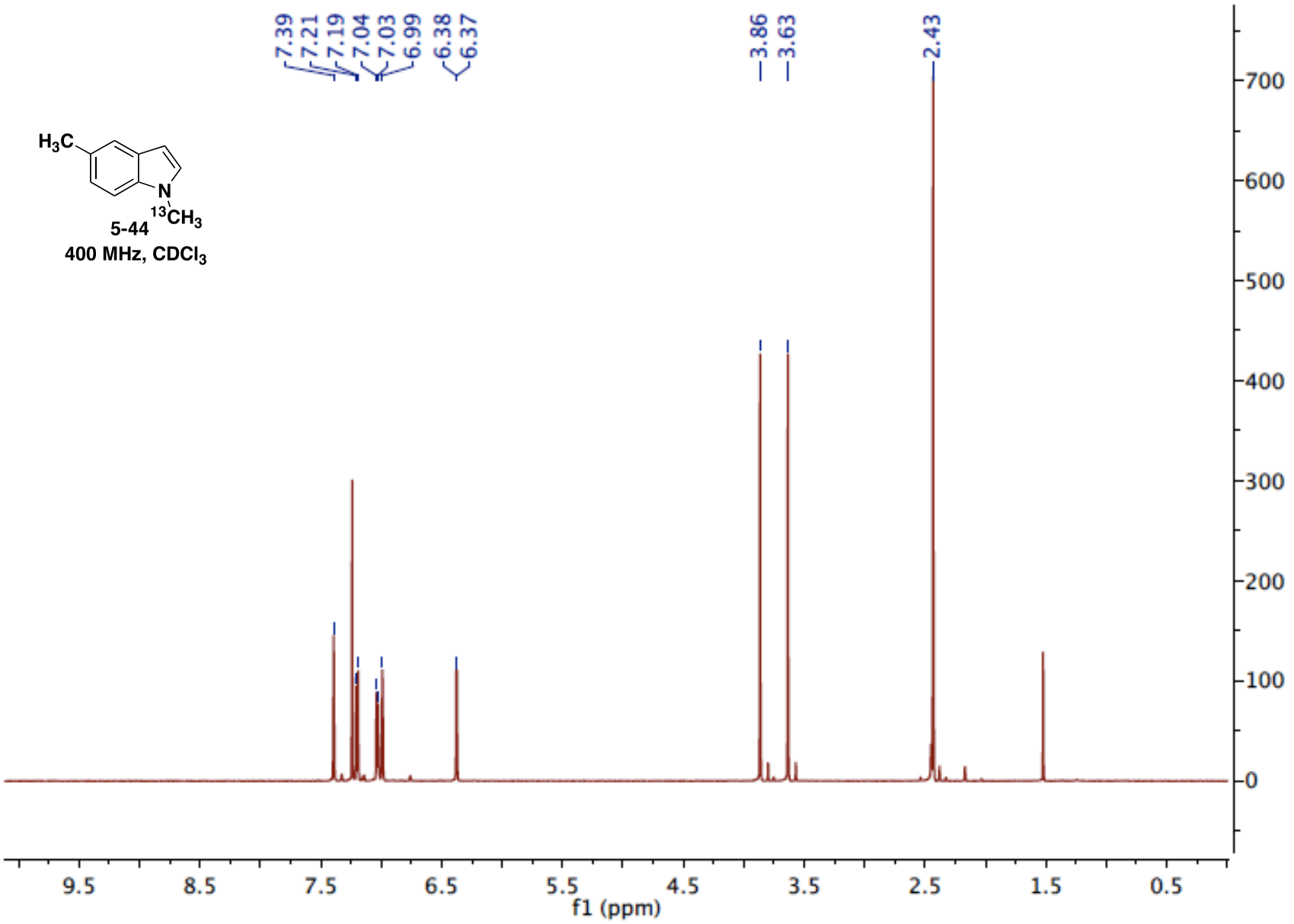

154.03
129.47
129.27
129.03
112.24
110.60
102.77
100.49
55.22
33.39
$^{13}CH_3$
N
OCH_3
5-41
100 MHz, $CDCl_3$
f1 (ppm)

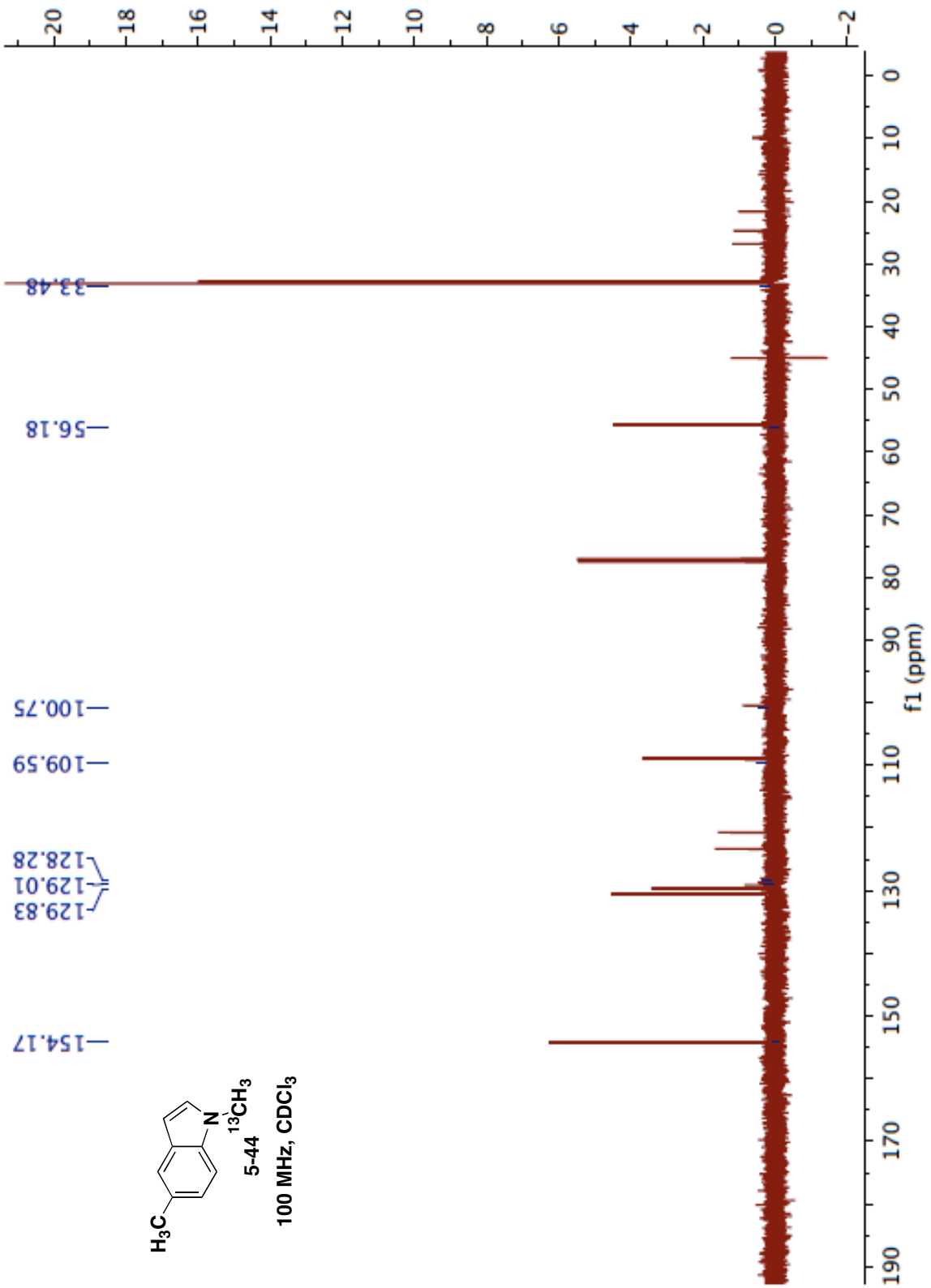

3.60
3.96
6.52
6.58
6.96
7.15
7
2
2
1
f1 (ppm)
5-42 $^{13}CH_3$
400 MHz, $CDCl_3$
H_3CO
N

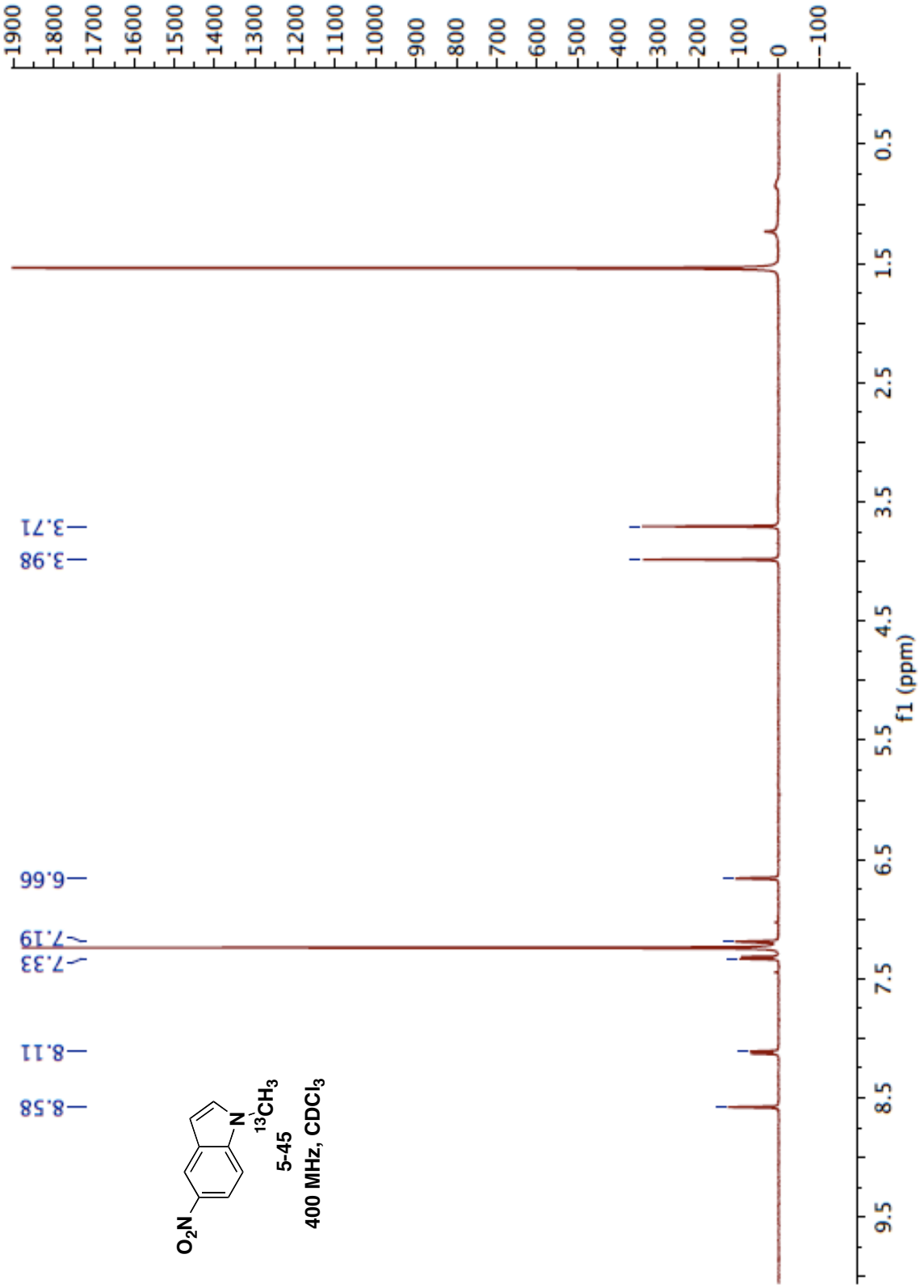

800
700
600
500
400
300
200
100
0
33.05
56.22
100.29
102.90
110.13
112.34
129.03
129.42
129.90
154.69
0
20
40
60
80
100
120
140
160
180
200
220
f1 (ppm)

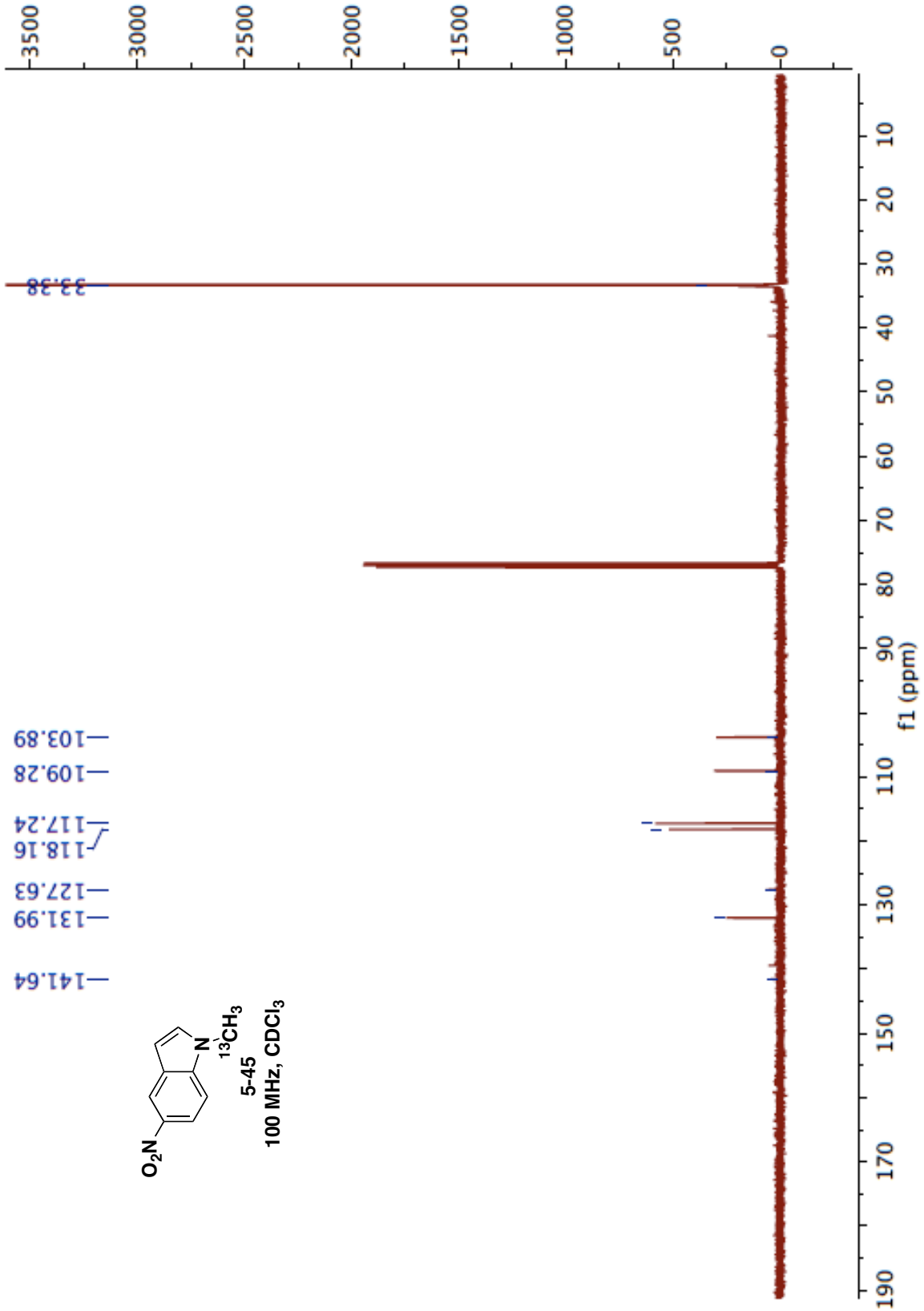

2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
−200
−2
−1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
f1 (ppm)
2.77
3.89
4.23
6.42
6.93
7.44
N
13CH3
CH3
5-43
400 MHz, CDCl3

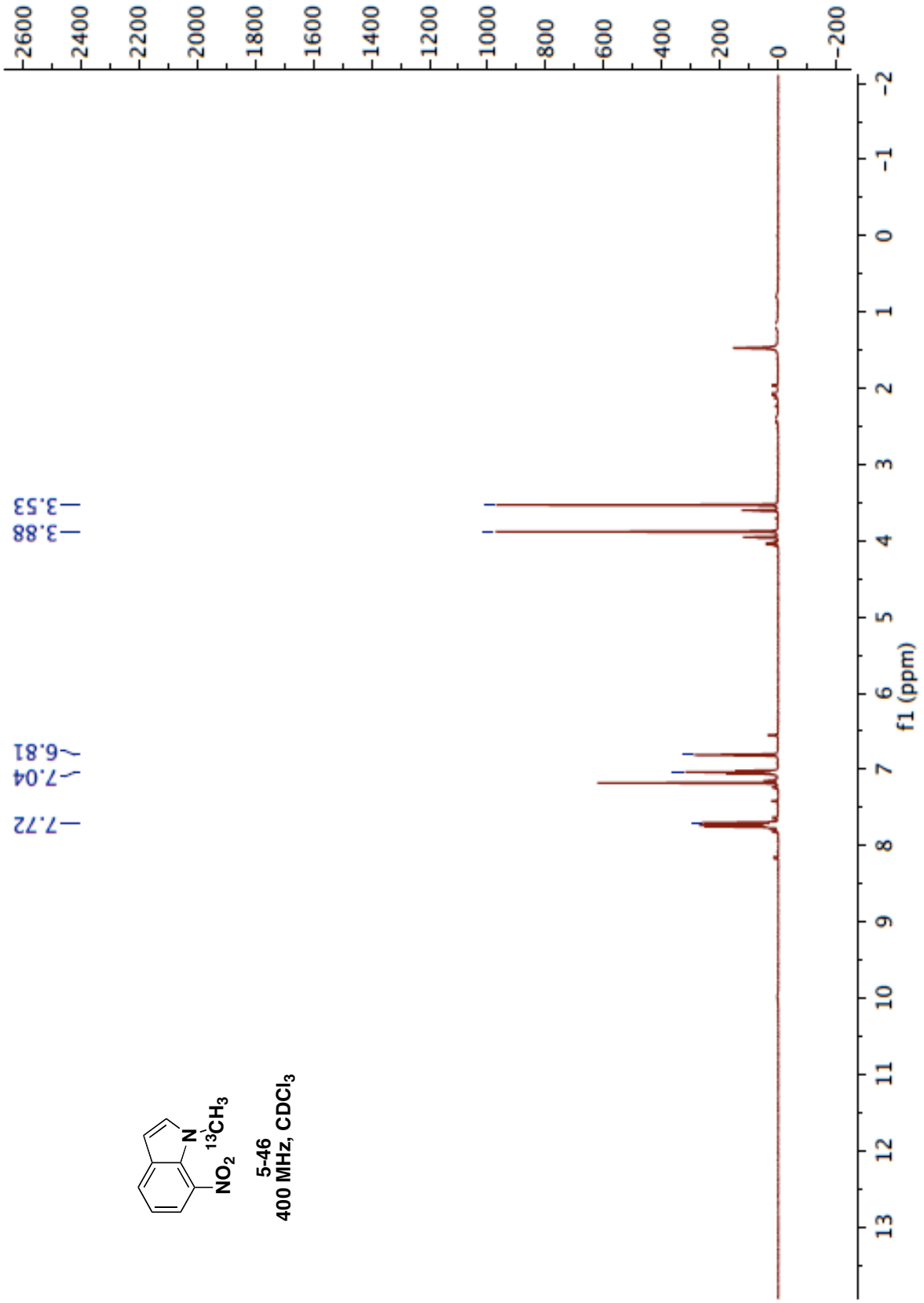

5-43
100 MHz, CDCl3
N
13CH3
CH3
19.96
37.00
101.10
101.08
119.28
119.75
121.36
124.32
129.81
130.55
f1 (ppm)

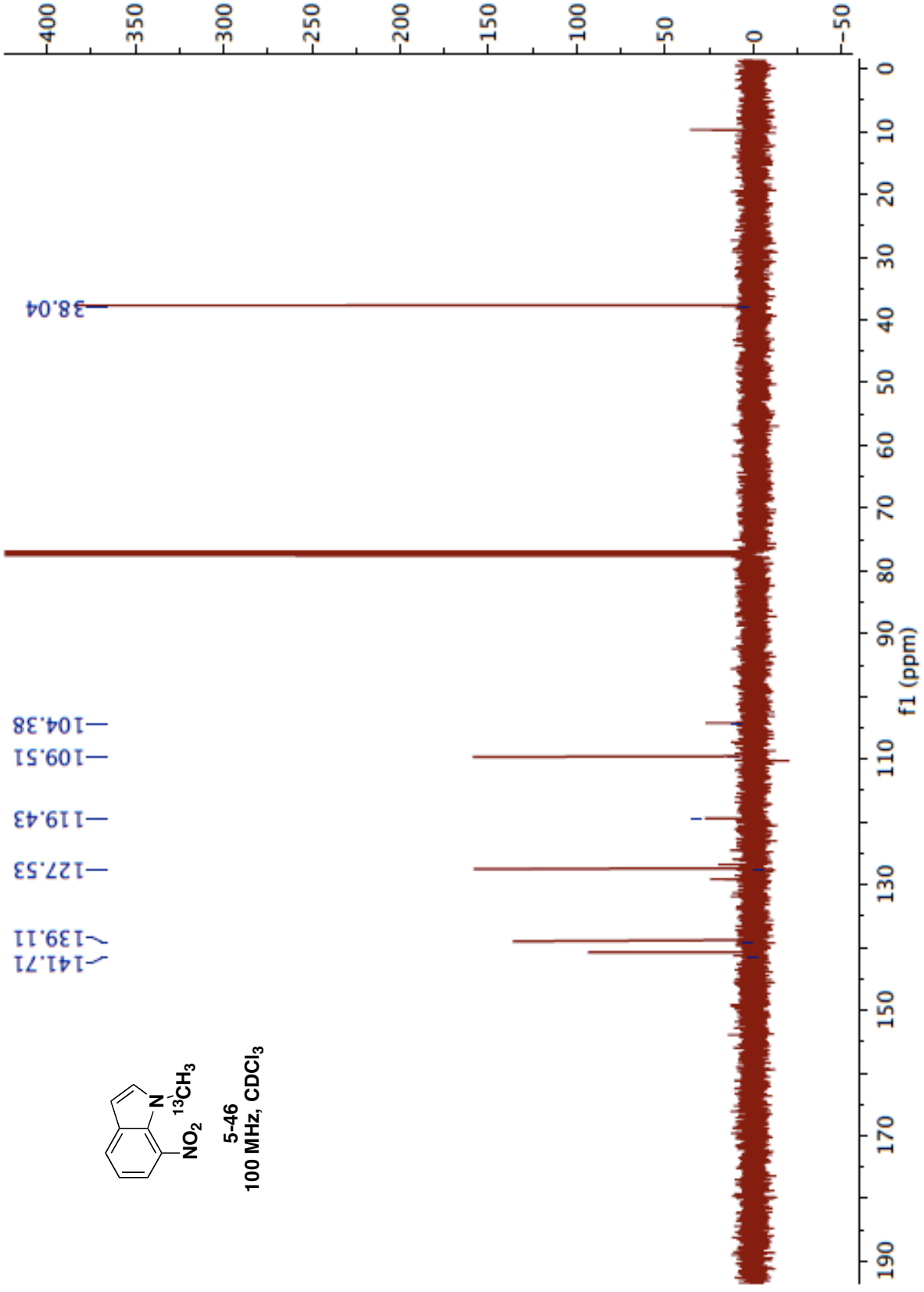

7.39
7.21
7.19
7.04
7.03
6.99
6.38
6.37
3.86
3.63
2.43
H3C
N
13CH3
5-44
400 MHz, CDCl3
700
600
500
400
300
200
100
0
9.5
8.5
7.5
6.5
5.5
4.5
3.5
2.5
1.5
0.5
f1 (ppm)

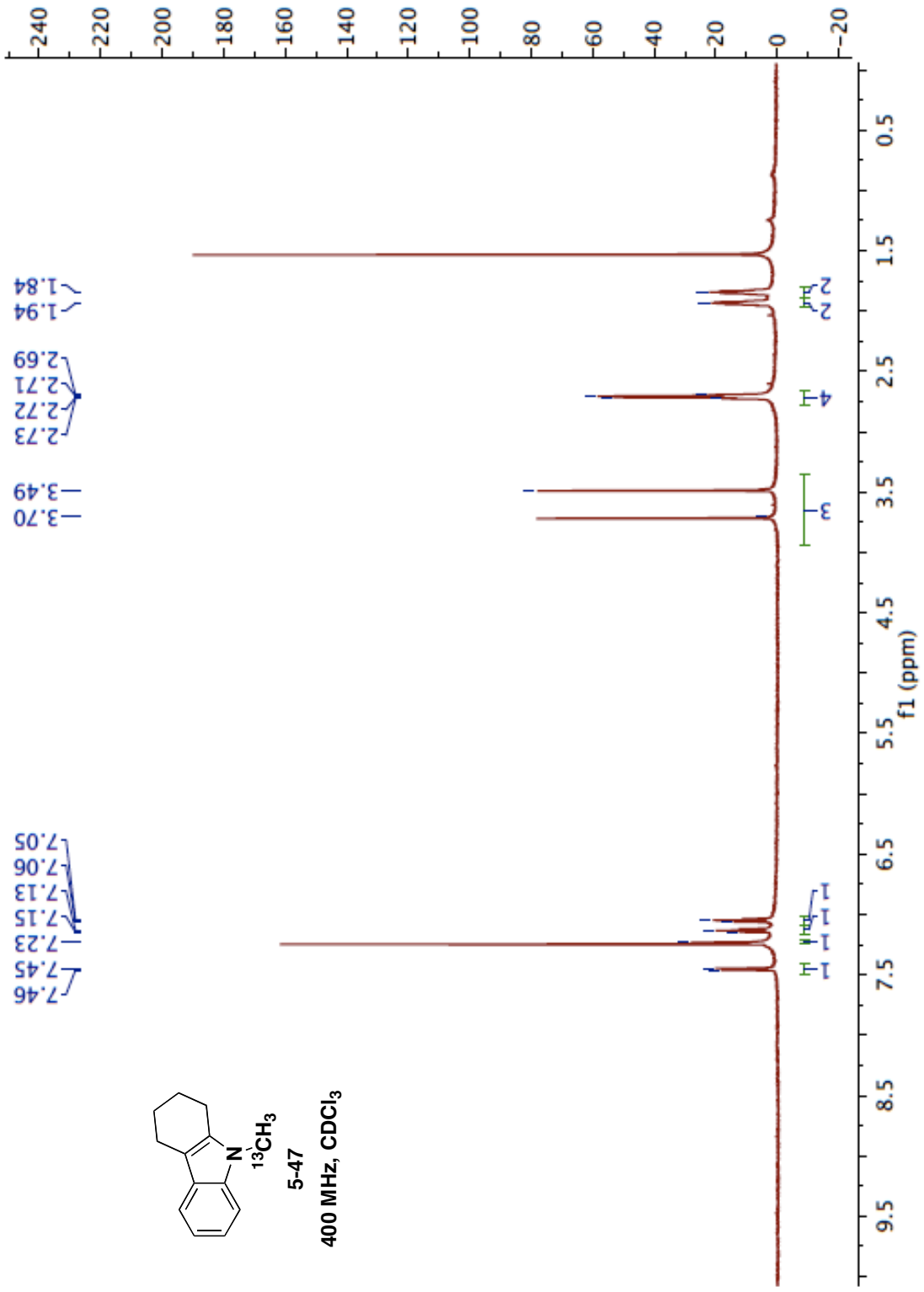

33.48
56.18
100.75
109.59
128.28
129.01
129.83
154.17
5-44
100 MHz, CDCl3
f1 (ppm)

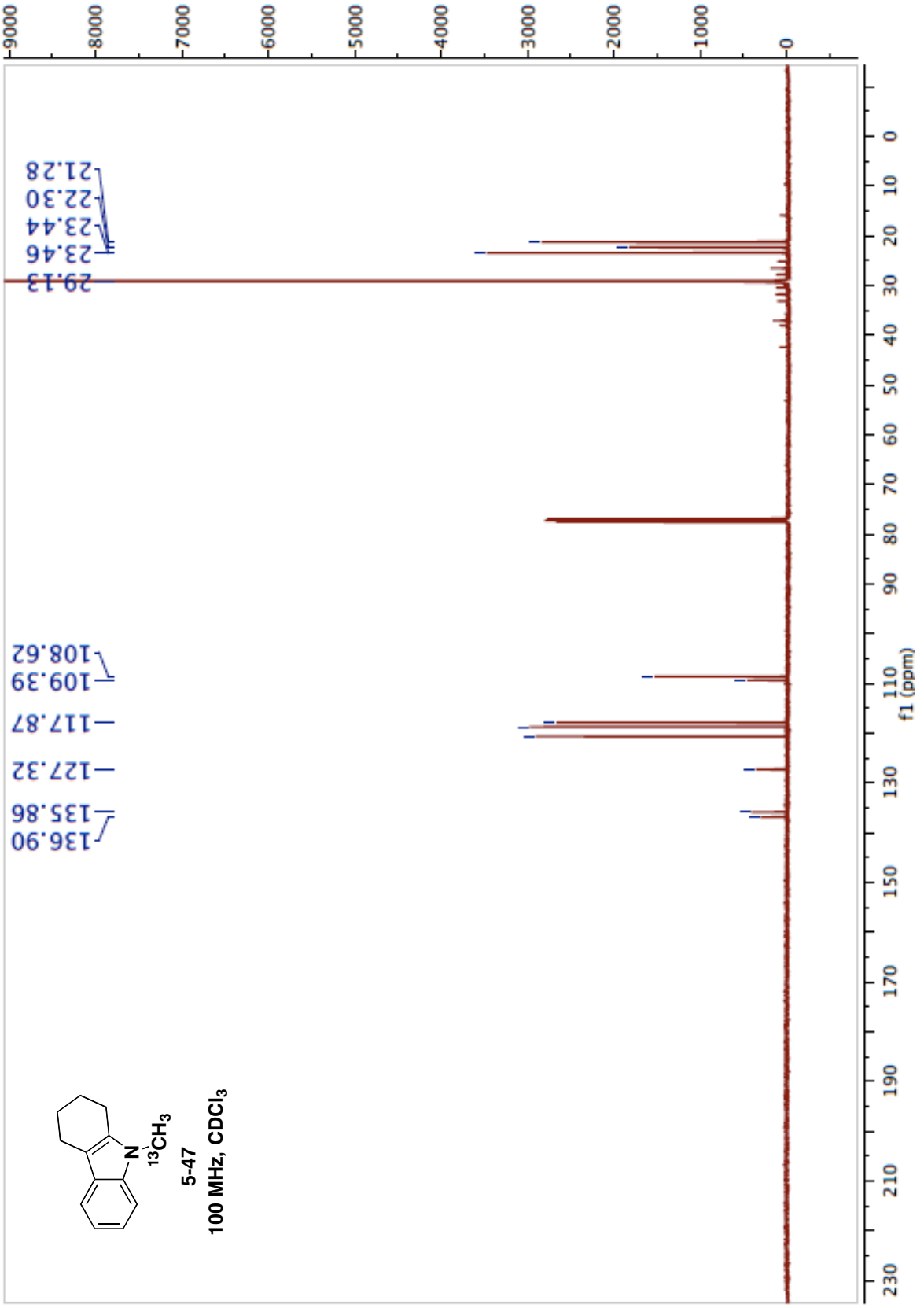

8.58
8.11
7.33
7.19
6.66
3.98
3.71
f1 (ppm)
5-45
400 MHz, CDCl3

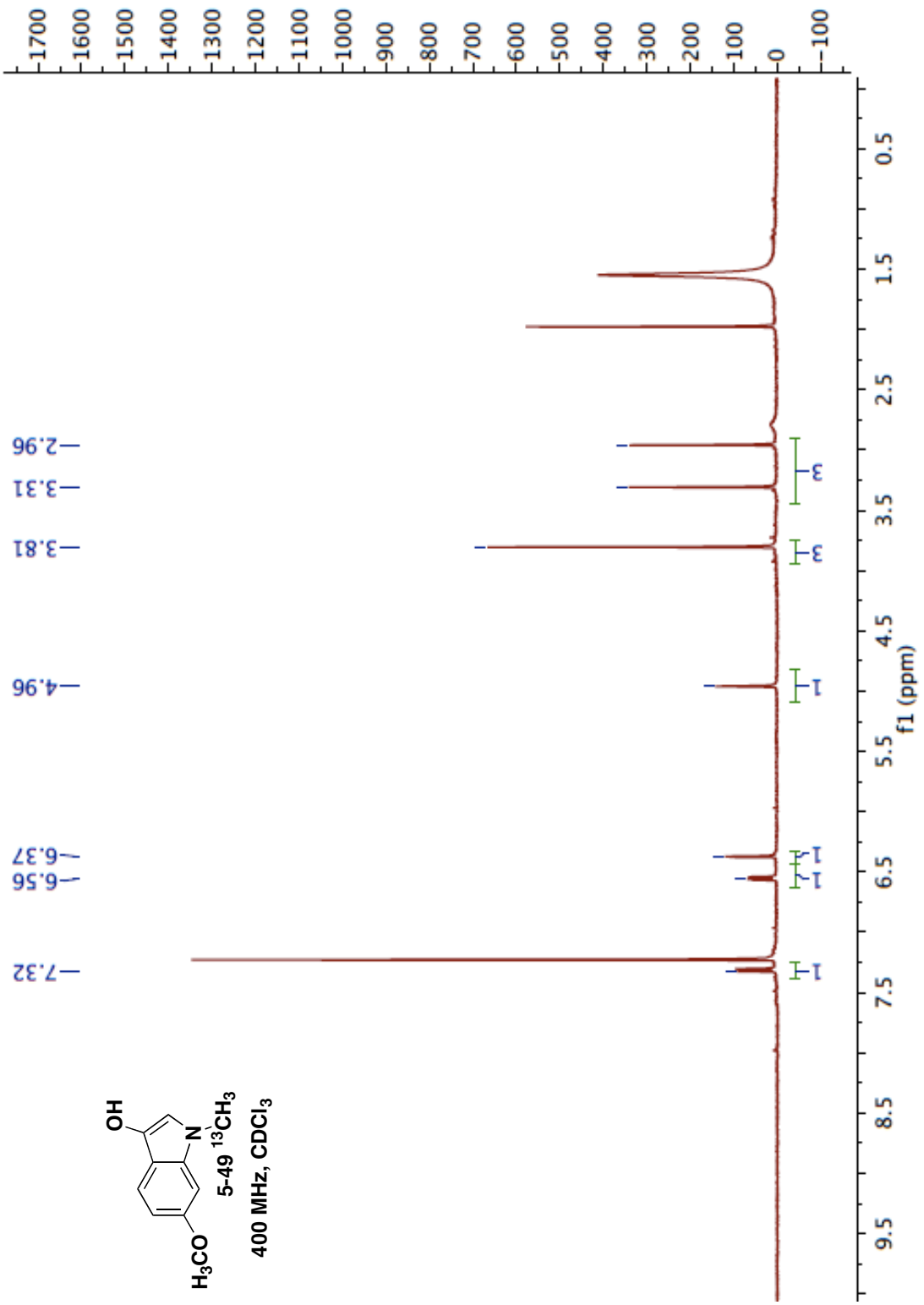

3500
3000
2500
2000
1500
1000
500
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
f1 (ppm)
33.38
103.89
109.28
117.24
118.16
127.63
131.99
141.64
O_2N
N
$^{13}CH_3$
5-45
100 MHz, $CDCl_3$

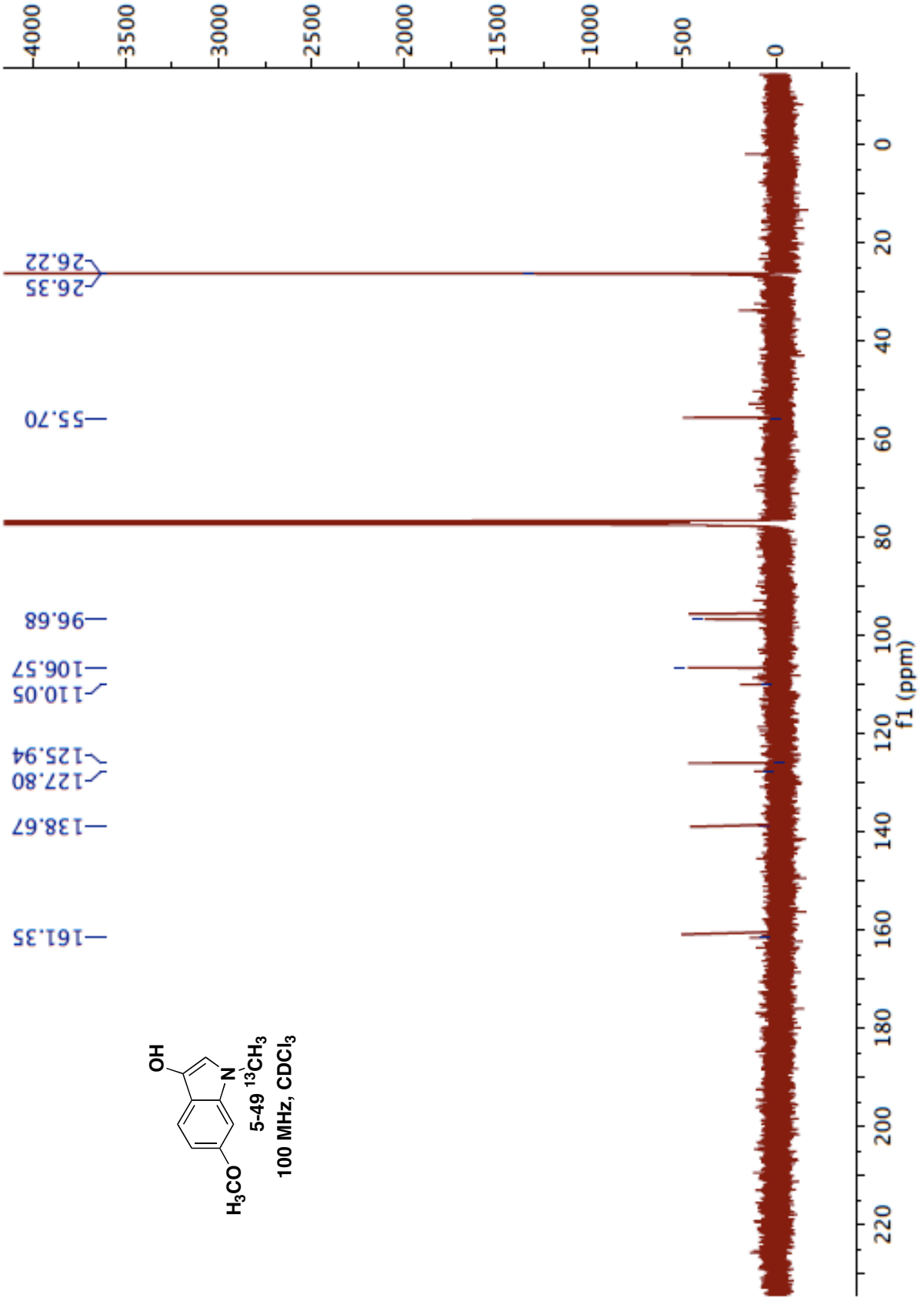

3.53
3.88
6.81
7.04
7.72
f1 (ppm)
5-46
400 MHz, CDCl3
N
13CH3
NO2

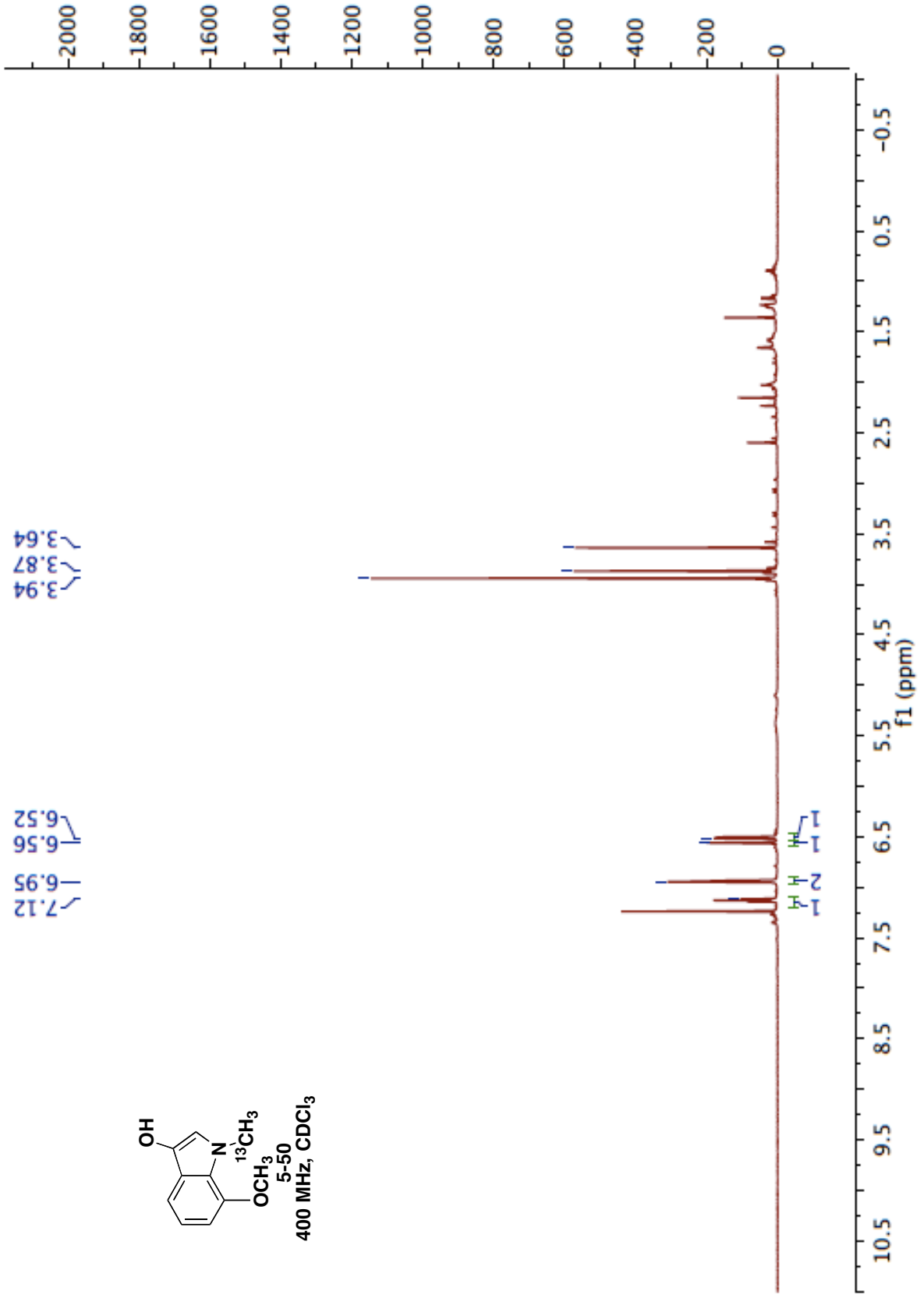

38.04
104.38
109.51
119.43
127.53
139.11
141.71
f1 (ppm)
5-46
100 MHz, CDCl3
N
13CH3
NO2

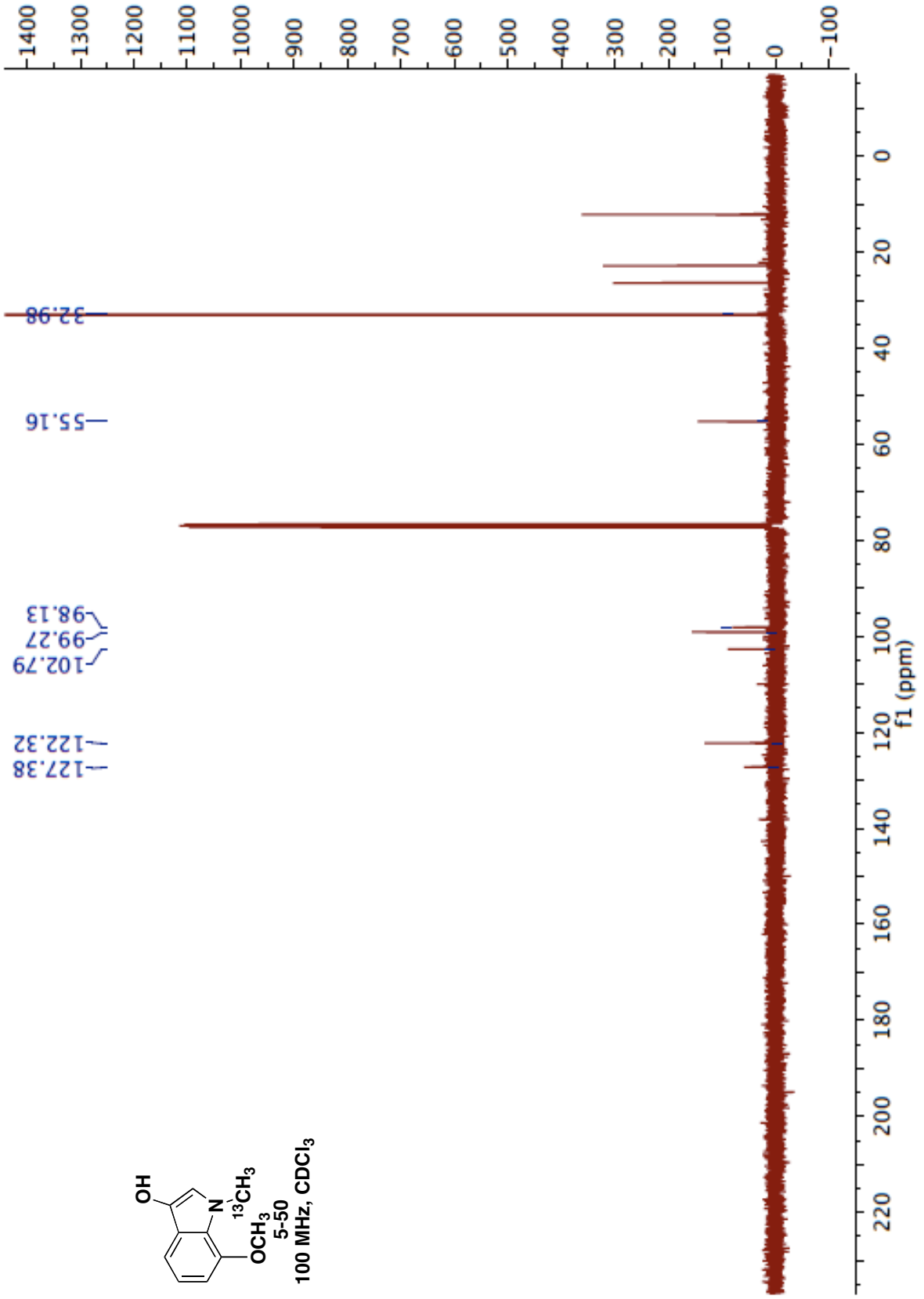

7.46
7.45
7.23
7.15
7.13
7.06
7.05
3.70
3.49
2.73
2.72
2.71
2.69
1.94
1.84
f1 (ppm)
5-47
400 MHz, CDCl3

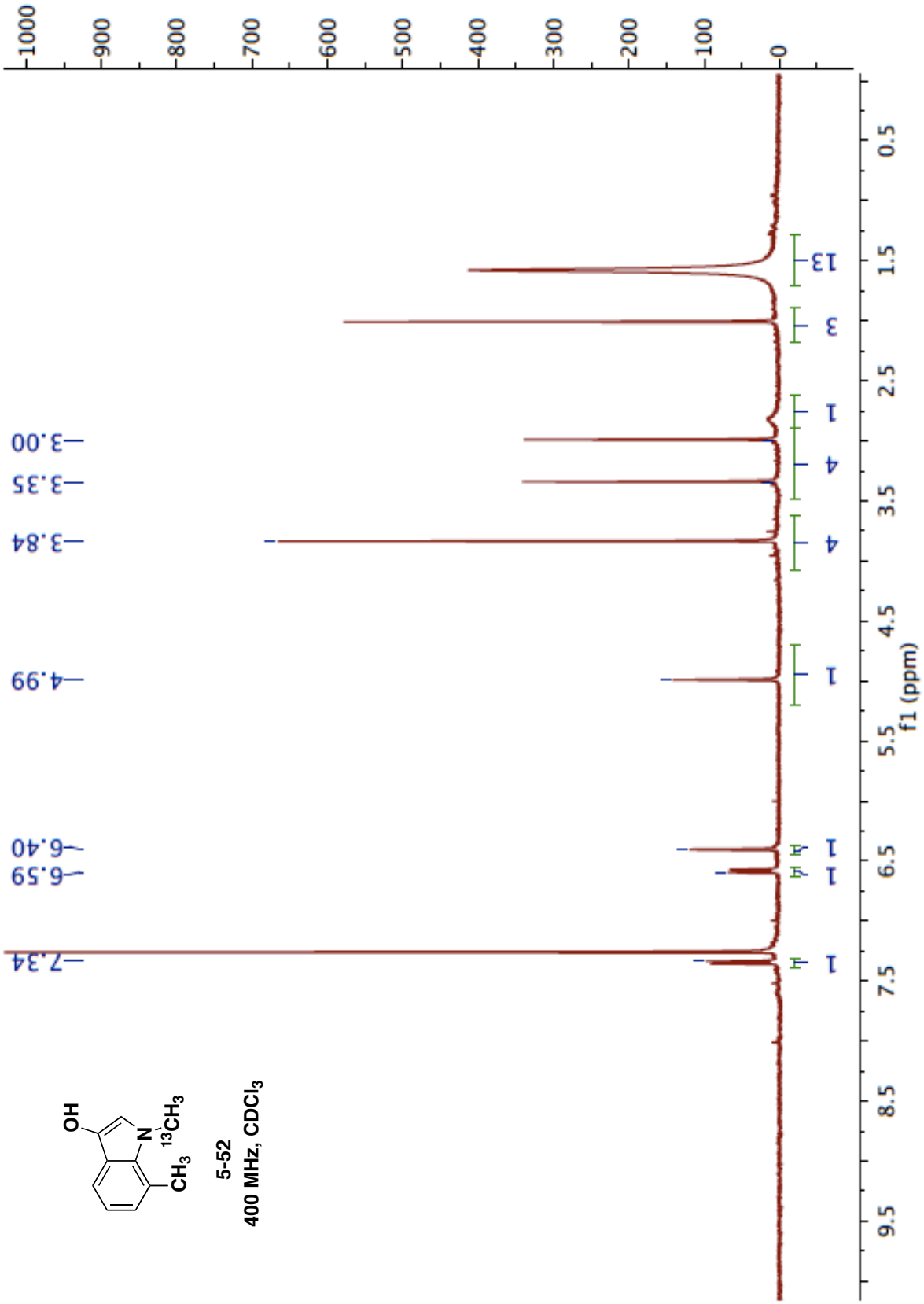

21.28
22.30
23.44
23.46
29.13
108.62
109.39
117.87
127.32
135.86
136.90
f1 (ppm)
5-47
100 MHz, CDCl3

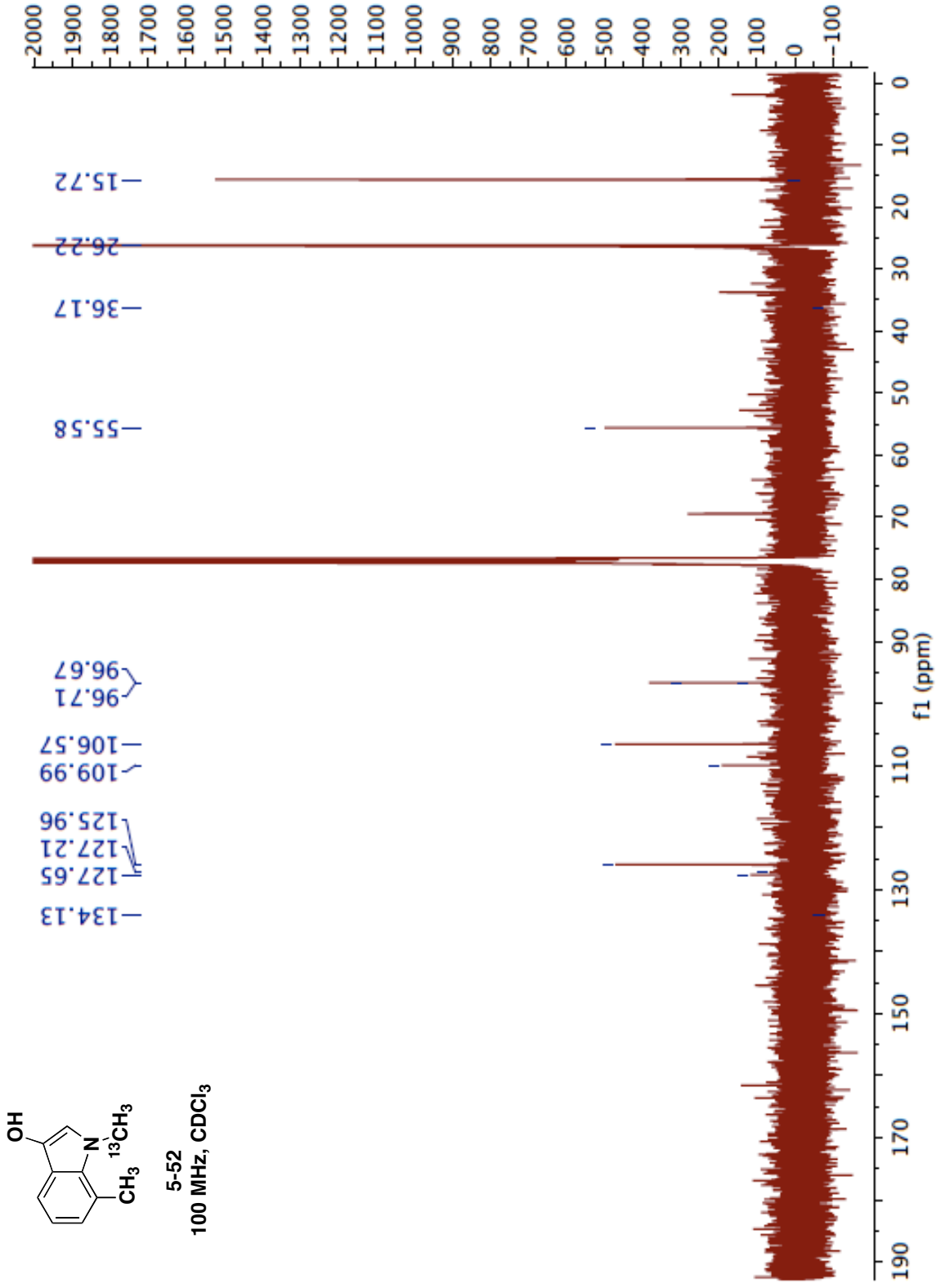

OH
N
$^{13}CH_3$
H_3CO
5-49
400 MHz, $CDCl_3$
7.32
6.56
6.37
4.96
3.81
3.31
2.96
f1 (ppm)

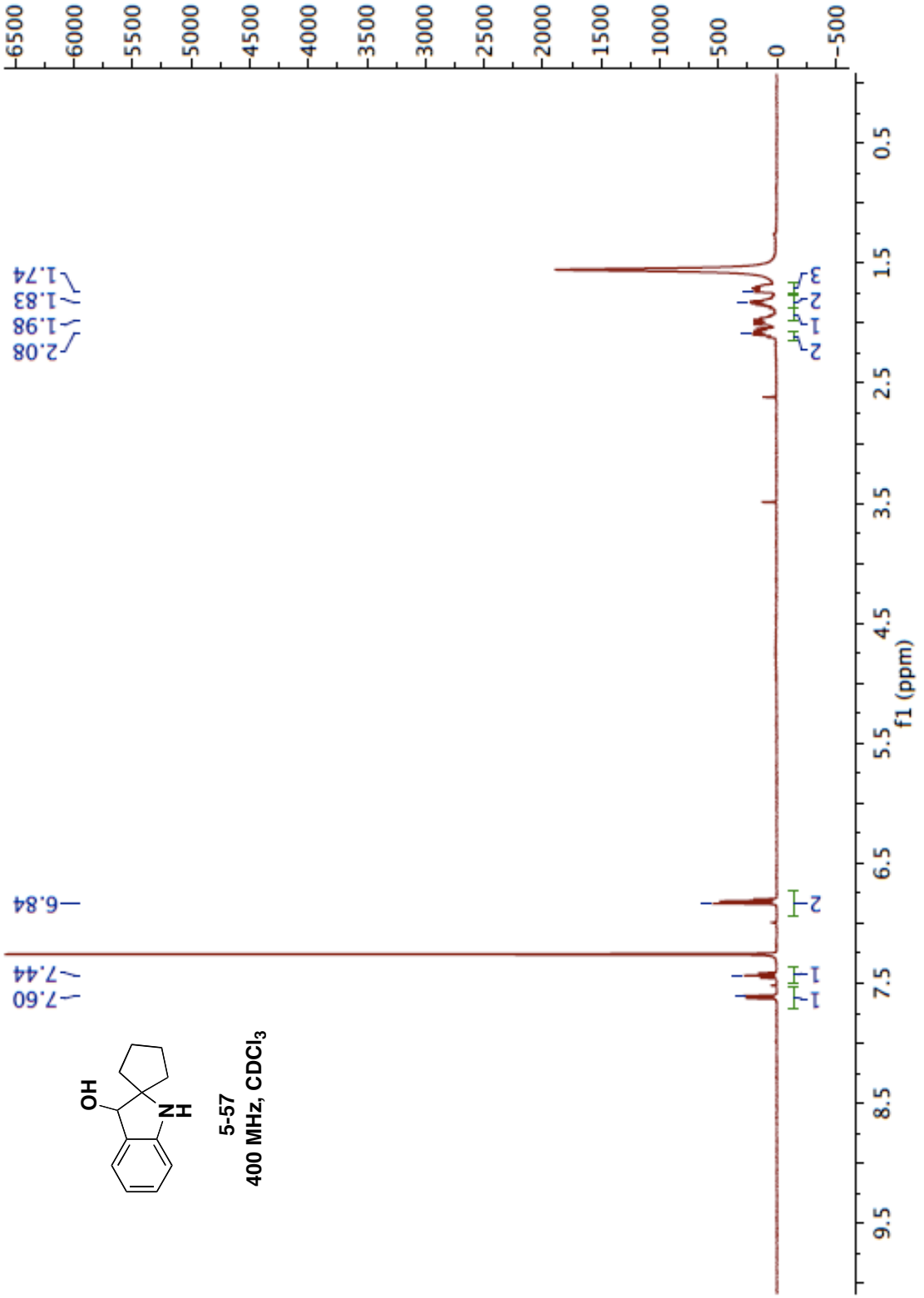

4000
3500
3000
2500
2000
1500
1000
500
0
26.22
26.35
55.70
96.68
106.57
110.05
125.94
127.80
138.67
161.35
0
20
40
60
80
100
120
140
160
180
200
220
f1 (ppm)
OH
N
$^{13}CH_3$
H_3CO
5-49
100 MHz, $CDCl_3$

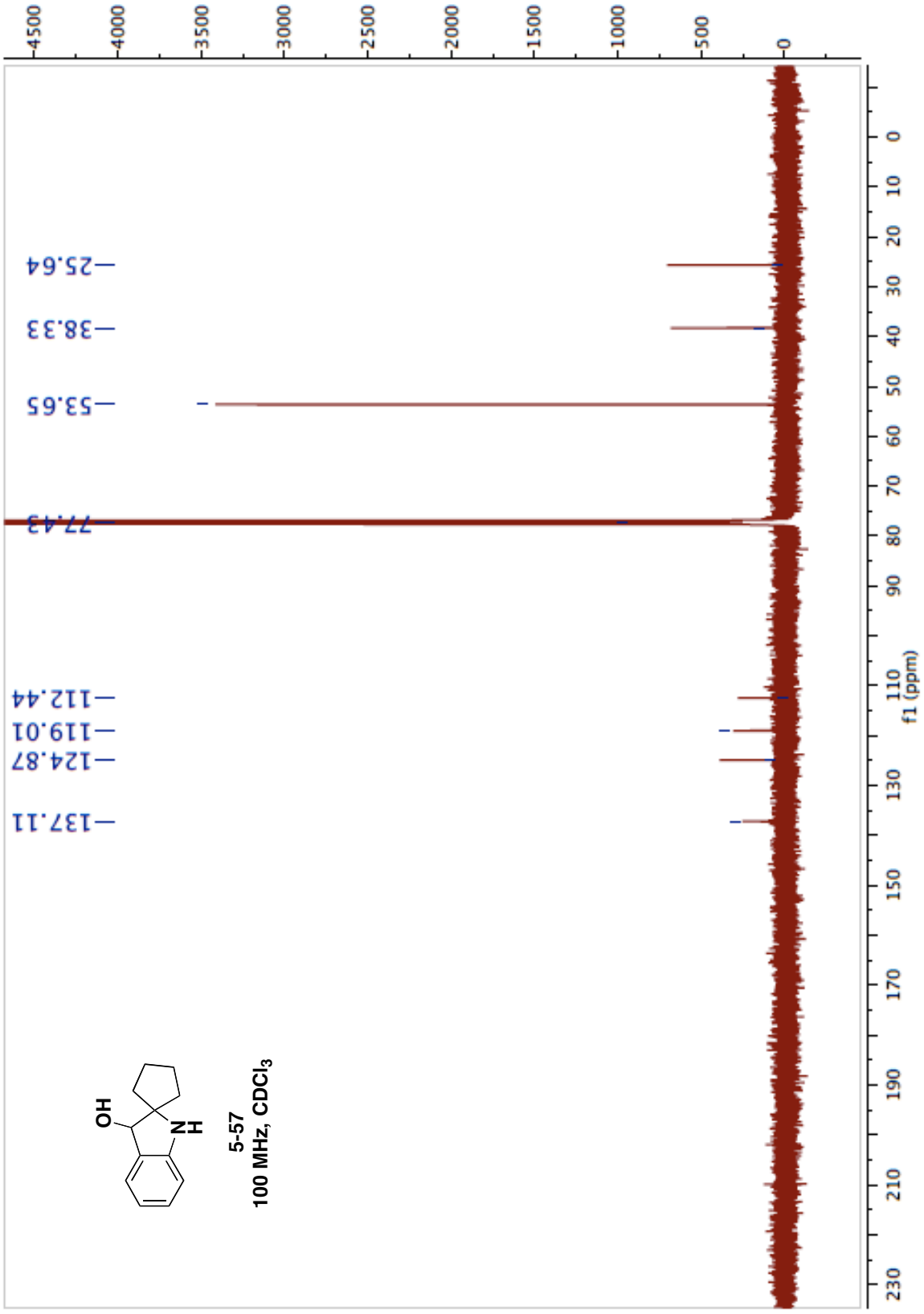

OH
N
$^{13}CH_3$
OCH_3
5-50
400 MHz, $CDCl_3$
3.64
3.87
3.94
6.52
6.56
6.95
7.12
f1 (ppm)

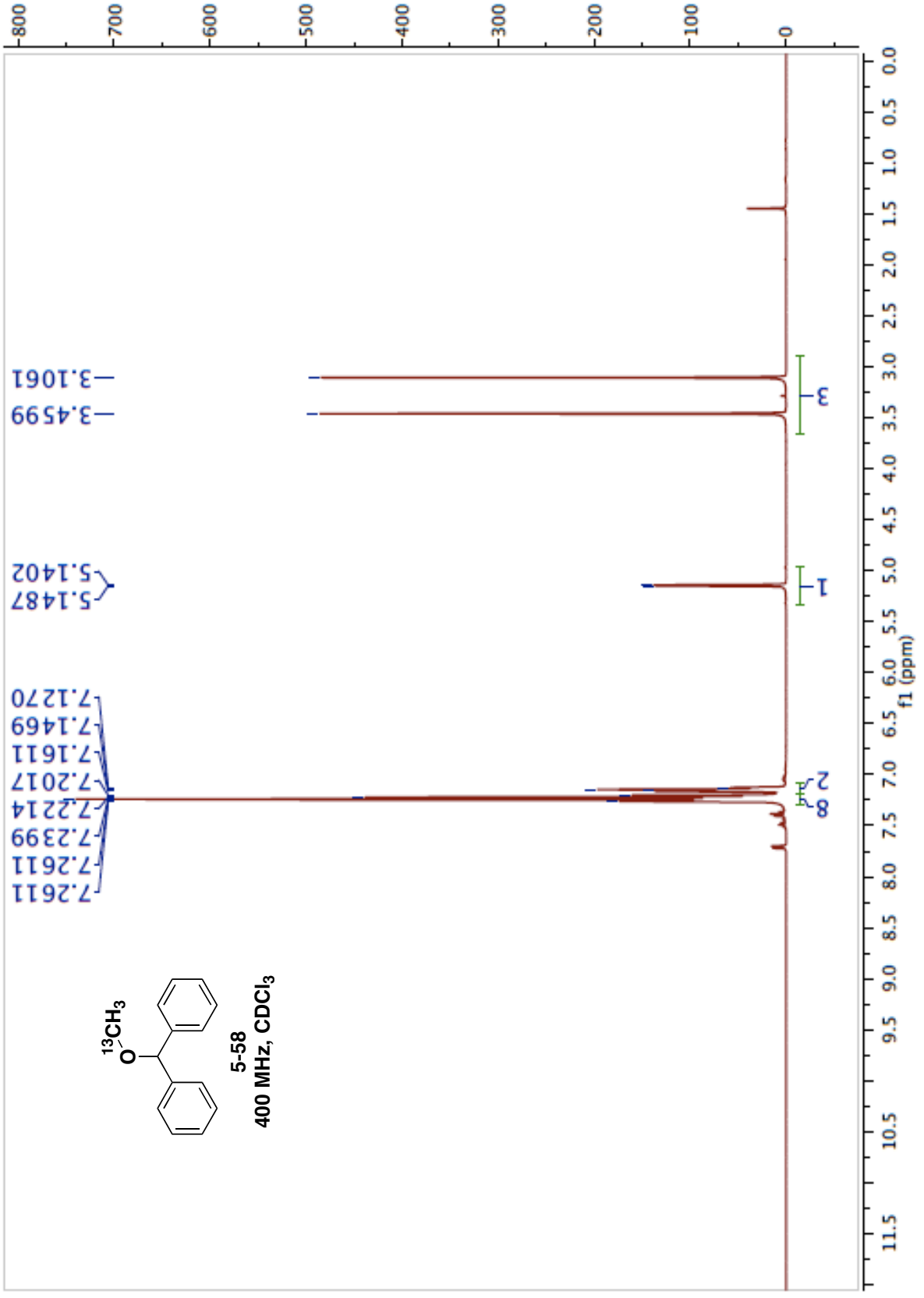

32.98
55.16
98.13
99.27
102.79
122.32
127.38
f1 (ppm)
OH
N
$^{13}CH_3$
OCH_3
5-50
100 MHz, $CDCl_3$

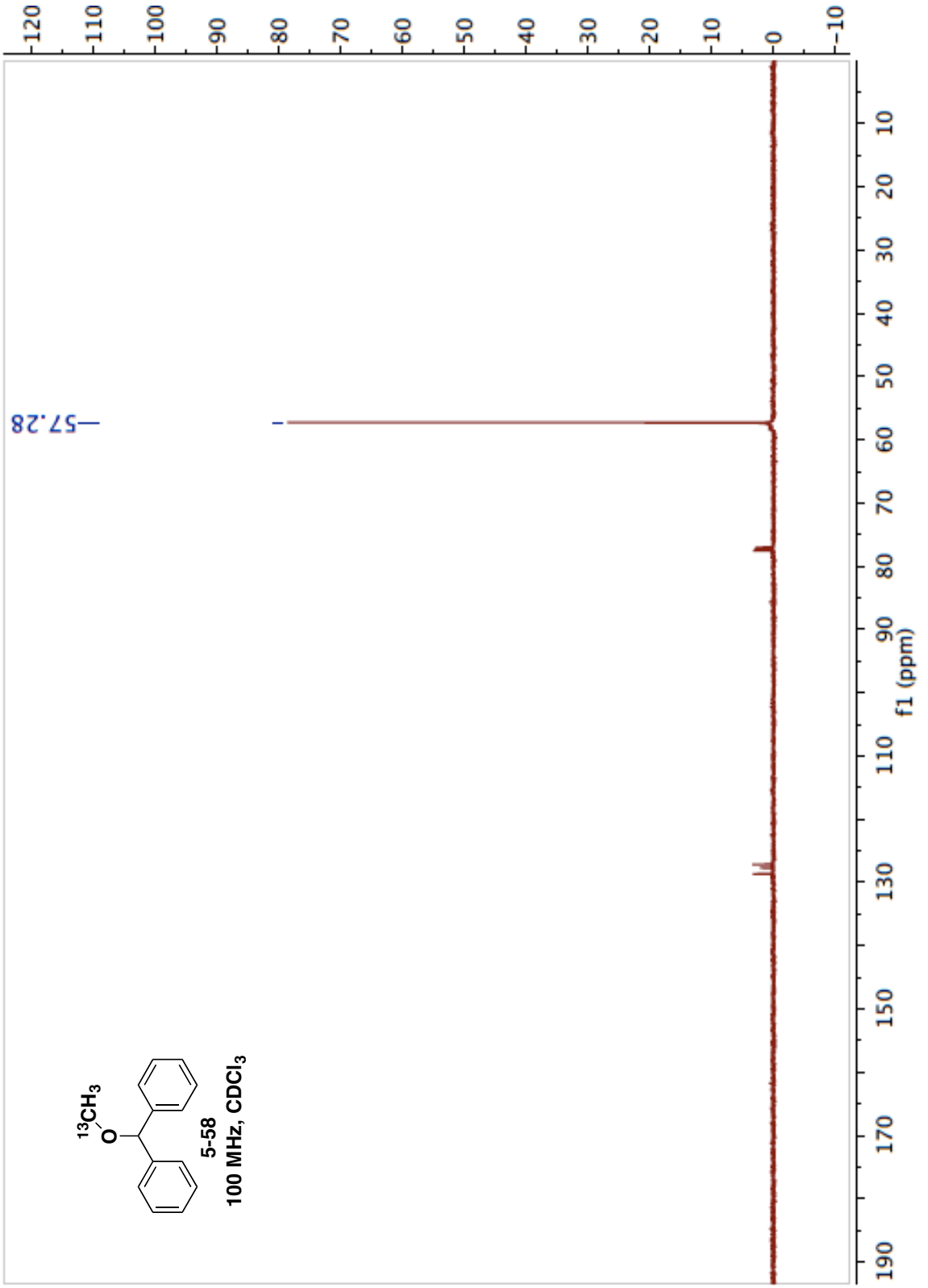

7.34
6.59
6.40
4.99
3.84
3.35
3.00
13
3
1
4
4
1
1
1
f1 (ppm)
OH
N
13CH3
CH3
5-52
400 MHz, CDCl3

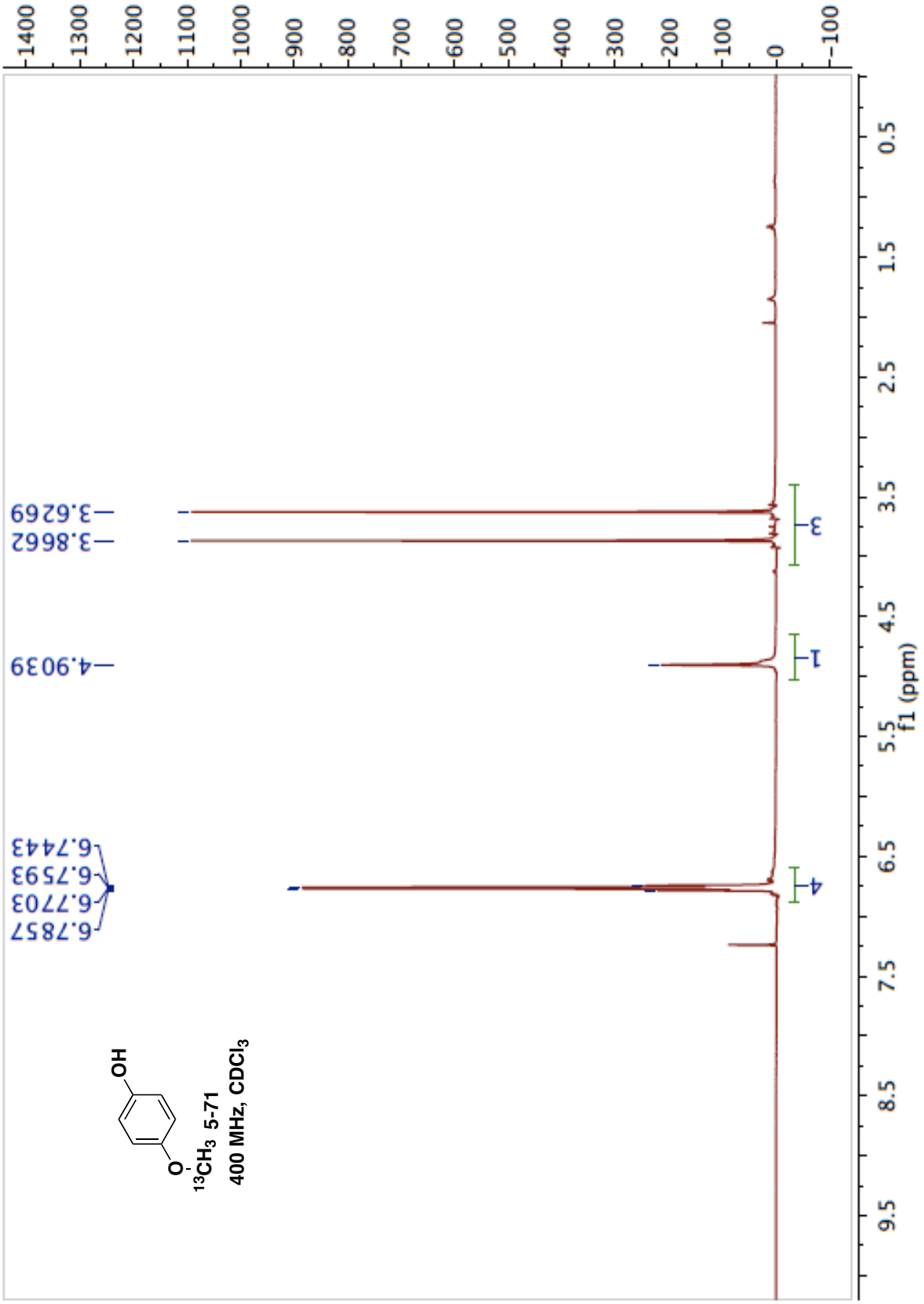

OH
N
$^{13}CH_3$
CH_3
5-52
100 MHz, $CDCl_3$
15.72
26.22
36.17
55.58
96.71
96.67
106.57
109.99
125.96
127.21
127.65
134.13
f1 (ppm)

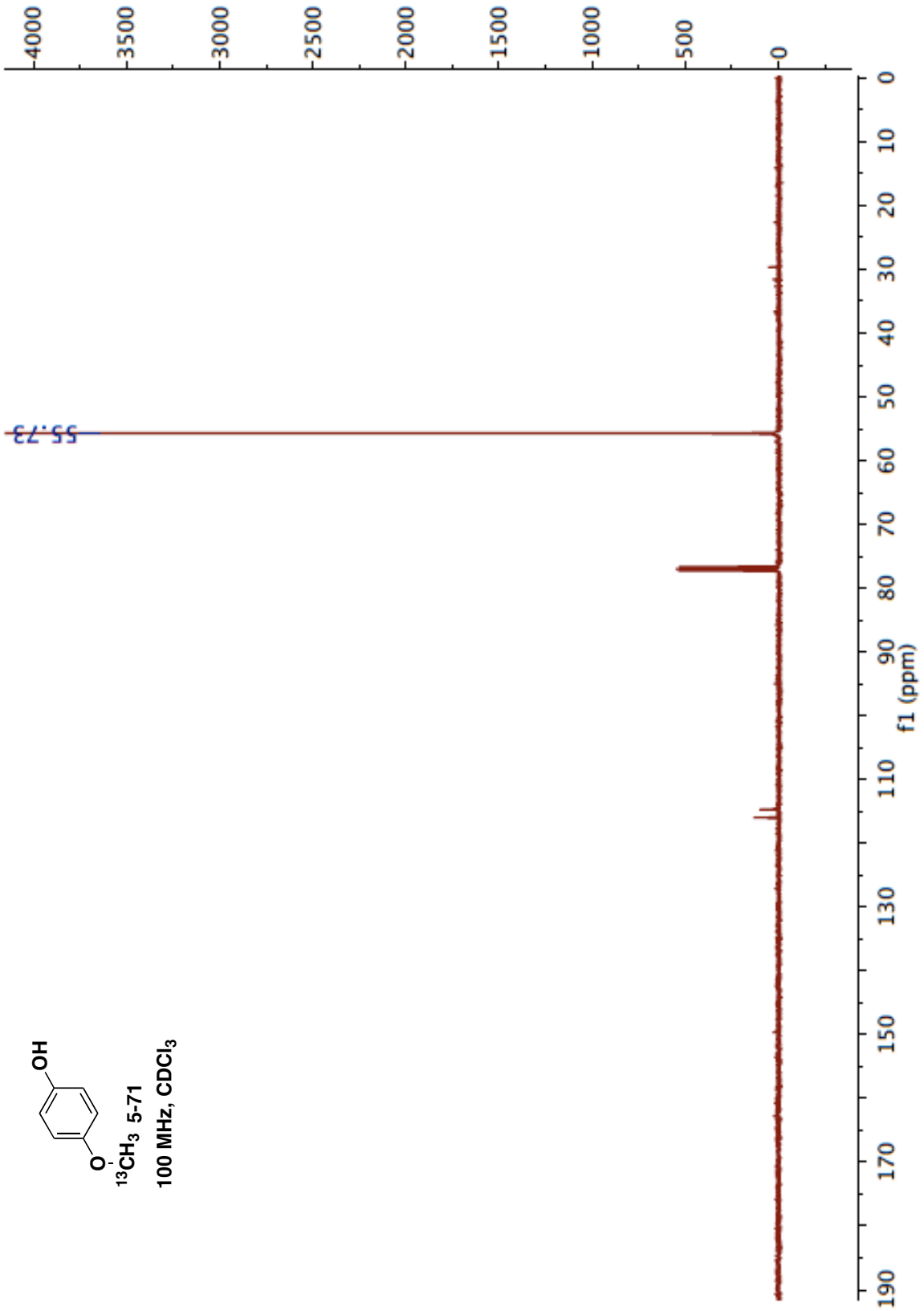

6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
−500
0.5
1.5
2.5
3.5
4.5
5.5
6.5
7.5
8.5
9.5
f1 (ppm)
1.74
1.83
1.98
2.08
6.84
7.44
7.60
OH
NH
5-57
400 MHz, CDCl3

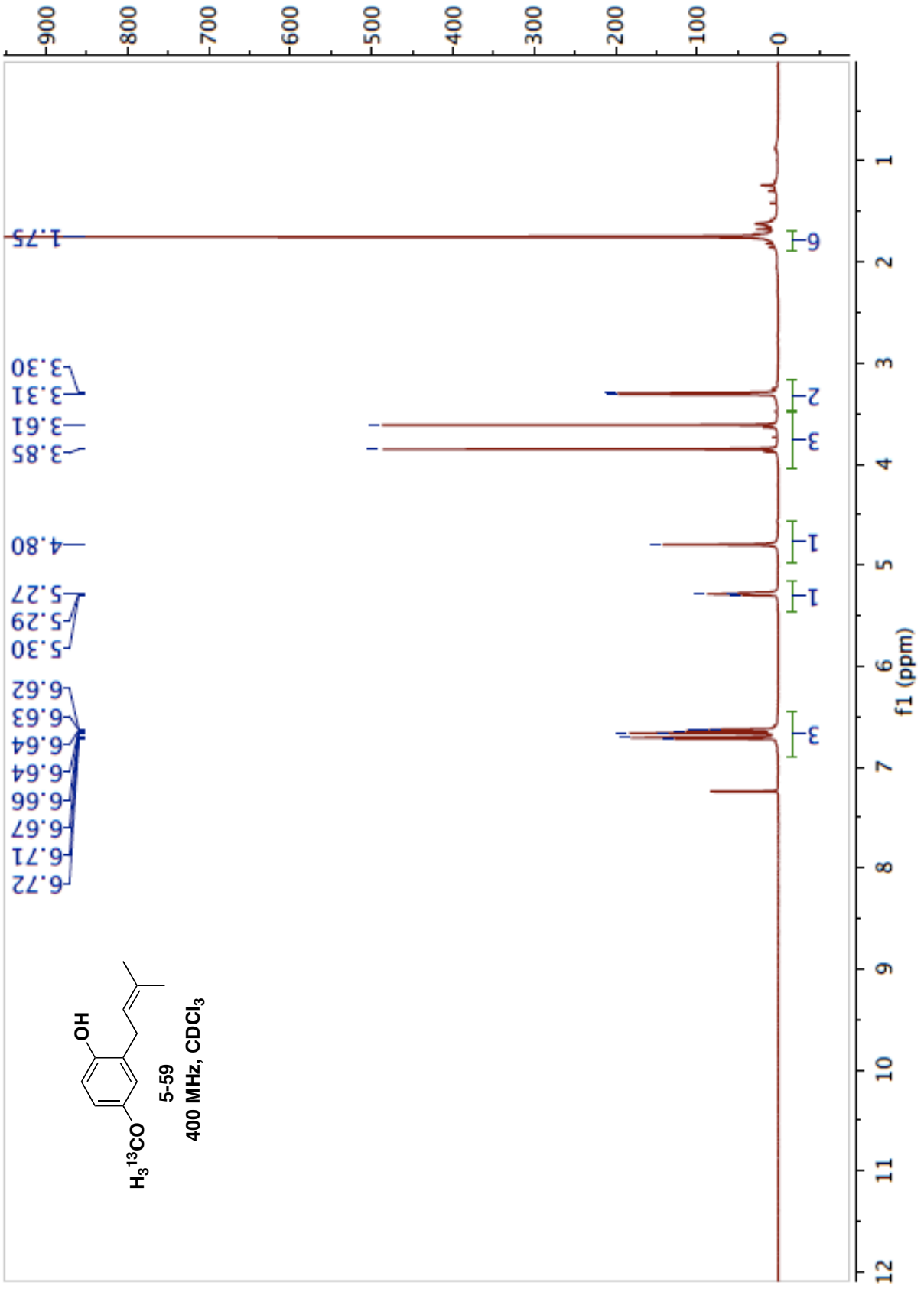

4500
4000
3500
3000
2500
2000
1500
1000
500
0
25.64
38.33
53.65
77.43
112.44
119.01
124.87
137.11
0
10
20
30
40
50
60
70
80
90
110
130
150
170
190
210
230
f1 (ppm)
OH
N
H
5-57
100 MHz, CDCl3

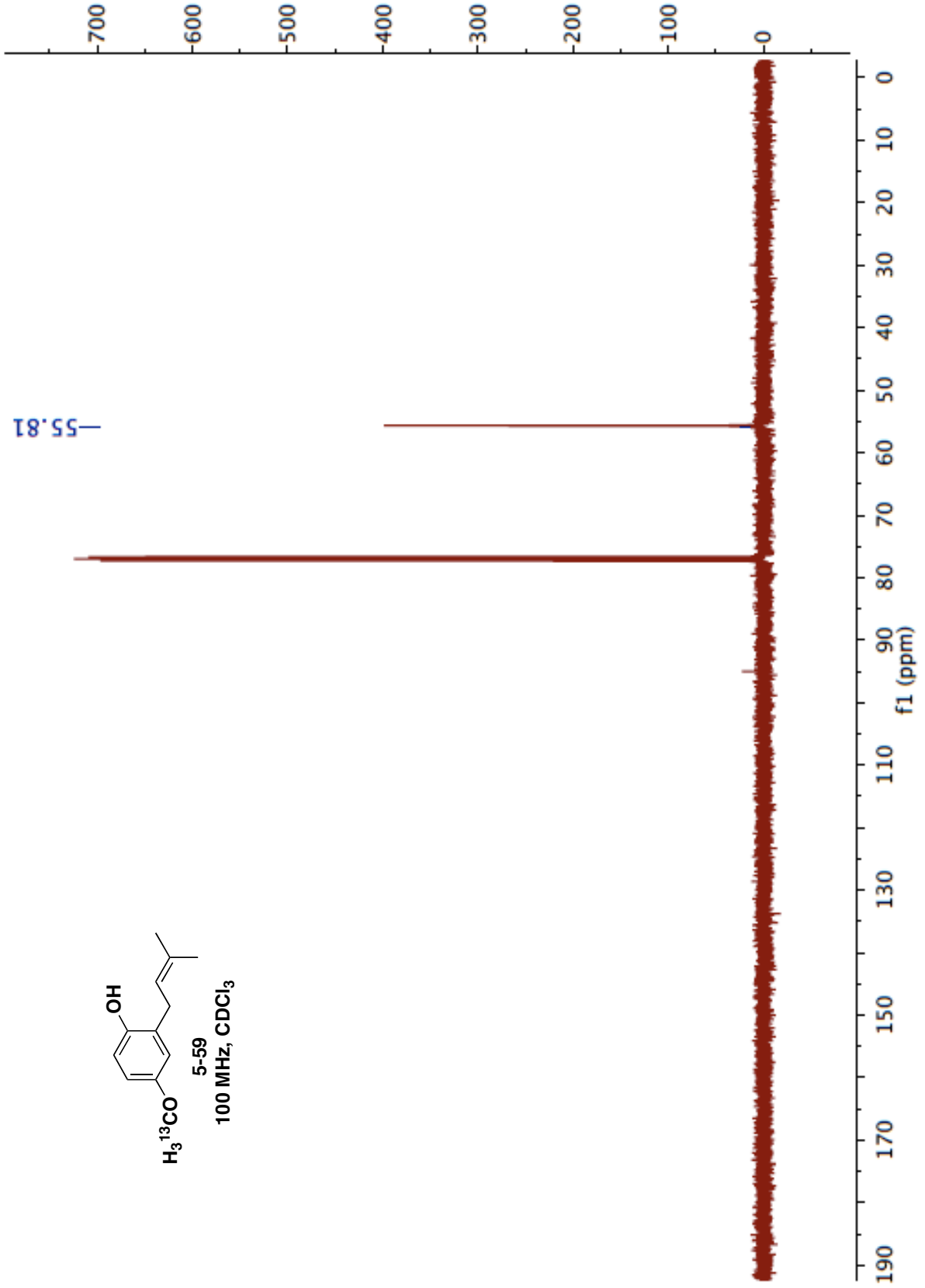

3.1061
3.4599
5.1487
5.1402
7.1270
7.1469
7.1611
7.2017
7.2214
7.2399
7.2611
7.2611
3
1
2
8
f1 (ppm)
$^{13}CH_3$
O
5-58
400 MHz, $CDCl_3$

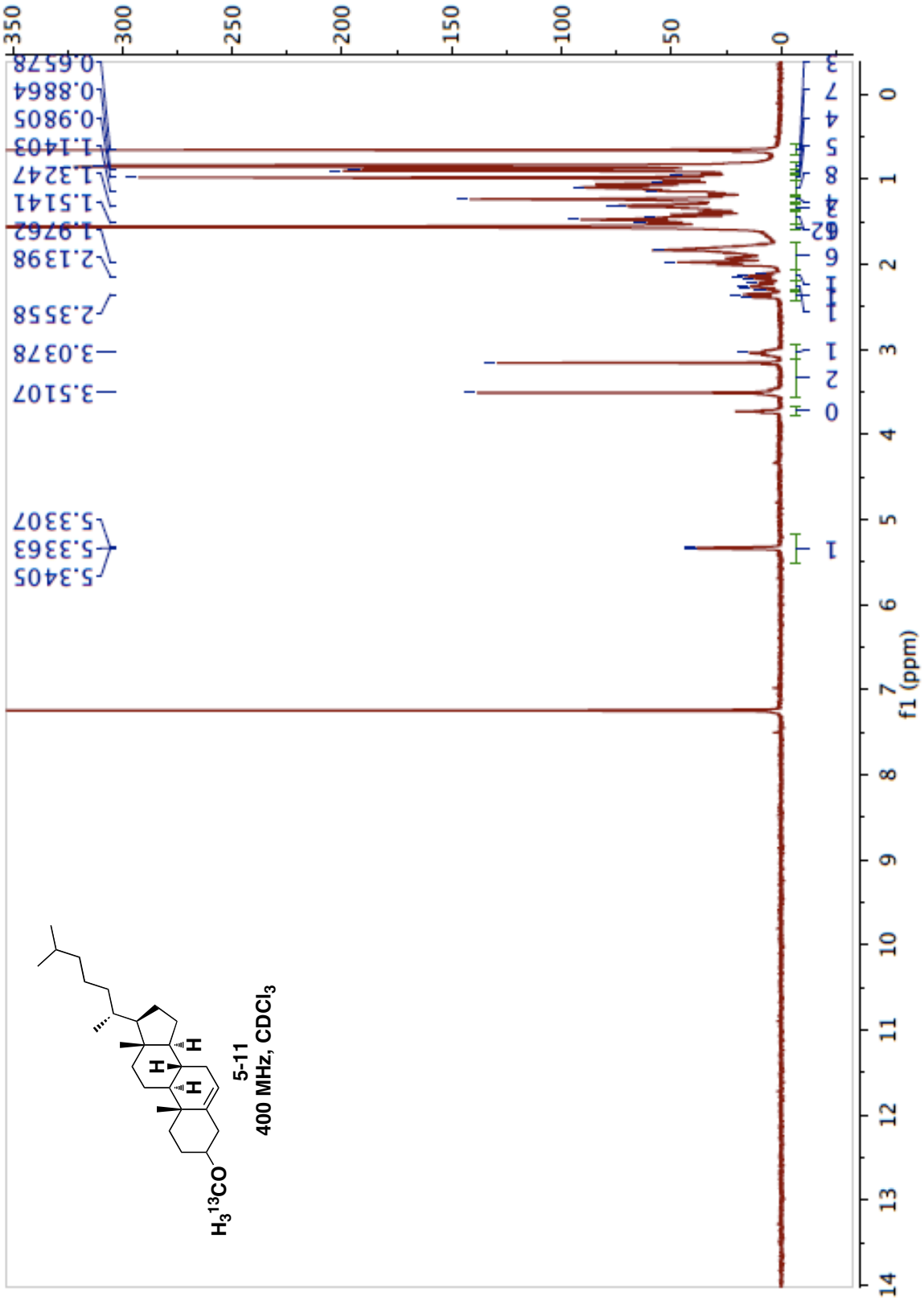

57.28
$^{13}CH_3$
O
5-58
100 MHz, $CDCl_3$
f1 (ppm)

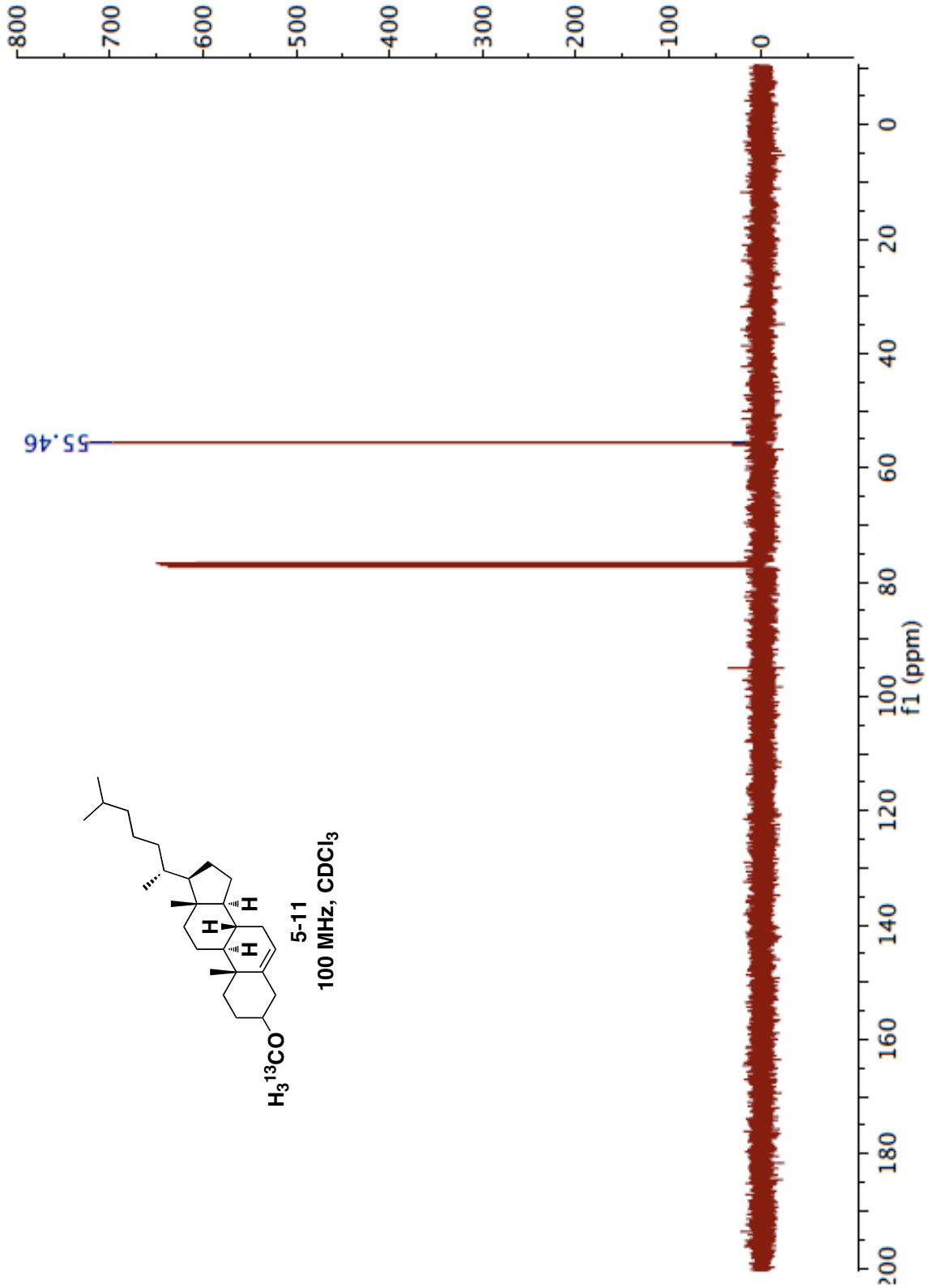

6.7857
6.7703
6.7593
6.7443
4.9039
3.8662
3.8269
4
1
3
f1 (ppm)
OH
O
$^{13}CH_3$ 5-71
400 MHz, $CDCl_3$


4000
3500
3000
2500
2000
1500
1000
500
0
55.73
0
10
20
30
40
50
60
70
80
90
110
130
150
170
190
f1 (ppm)
OH
O
$^{13}CH_3$ 5-71
100 MHz, $CDCl_3$


1.75
3.30
3.31
3.61
3.85
4.80
5.27
5.29
5.30
6.62
6.63
6.64
6.64
6.66
6.67
6.71
6.72
6
2
3
1
1
3
f1 (ppm)
OH
$H_3{}^{13}CO$
5-59
400 MHz, $CDCl_3$


55.81
f1 (ppm)
OH
$H_3{}^{13}CO$
5-59
100 MHz, $CDCl_3$


5-11
400 MHz, CDCl3
H3 13CO
f1 (ppm)


55.46
f1 (ppm)
5-11
100 MHz, CDCl3
H3 13CO