Structural Elucidation of An Unknown Compound

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Yan Yu

August 2018

ABSTRACT

High resolution mass spectroscopy (HRMS) and nuclear magnetic resonance (NMR) are both powerful ways to identify the structure of an unknown molecule. They are widely used in determining natural products and other newly synthesized molecules. In this thesis, several types of NMR spectra and a HRMS data of an unknown molecule were provided, and based on those data, the complete structure of this molecule was identified.

BIOGRAPHICAL SKETCH

Yan Yu is an honor graduate (B.S.) with the highest extinction from Purdue University, West Lafayette, IN in May 2013 and will be a graduate (M.S.) from Cornell University, Ithaca, NY in August 2018. He started his research interest in chemical biology during his year at Purdue University with Prof. Jean Chmielewski and Prof. Alexander Wei, who influenced and motivated him to pursue an advanced degree in the same field. At Cornell, he took advanced classes and gained great inspiration from several professors, including Prof. Yimon Aye and Prof. Frank Schreoder, and decided to pursue a Ph.D. degree in the field of chemical biology since August 2018.

Acknowledgements

I would like to thank my family, who has always motivated me to explore more and learning. Their encouragement strongly influenced my decisions to attend graduate school, pursue an advanced degree, and professionalize myself to become a true chemist.

I would also like to thank all the professors who mentored and supported me to this point in my education. Without their patience, encouragements, and guidance, I would have never pursued my career in science.

TABLE OF CONTENTS

2. NMR & MS Experiments: Results

LIST OF FIGURES

Page
Number

LIST OF TABLES

Chapter 1 Introduction

Structural determination of an unknown molecule is one of the most important efforts in identifying natural products. Historically, the discovery of palytoxin[1] played an important role in both chemistry and pharmacology, so does many other molecules. For structure elucidation, people often use mass spectrometry[2] and NMR spectroscopy[3].

Mass spectrometry is an analytical method that ionizes molecules and measure their mass-to-charge ratio[4]. Electrospray ionization (ESI) is a common "soft" ionization method, and for small molecules, it often produces single-charged ions, which makes it easy to determine the molar mass[5]. Depending on the structure of the molecule, ions generated from ESI can carry either positive charge by protonation, or negative charge by deprotonation. ESI coupled with high-resolution mass spectrometry provides both high resolution and high mass accuracy, and it is a powerful tool for the analysis of elemental compositions and identification of unknowns[6]. The isotope pattern will show up on the spectrum and help determine the existence of heteroatom[7]. Compared to low resolution mass spectrometry, where the resolution is one mass unit, HRMS has a resolution that is 0.00001 mass unit, so the exact mass of the molecule can be obtained. Because different elements, although some are similar in mass, have different exact mass, there are only a few combinations that sum up to the exact mass of the molecule, which means that by going through a library of exact atomic mass, the information of elemental composition can be gained. In this thesis, ESI-HRMS in negative mode was used to determine the exact molecular weight, which was then used to determine the molecular formula.

NMR spectroscopy is another analytical tool that detects and distinguishes a specific atomic nucleus and characterizes its local chemical environment, based on the difference of the local magnetic field—which is a sum of the applied magnetic field and local variations due to the surrounding electronic environment—in the vicinity of those atomic nuclei[8]. It excites nucleus by changing the nucleus spin and measure the relaxation time. The resulting spectrum can be used for structural elucidation of the molecule[9]. Historically, an early application of NMR was the structure of C.101 – C.115 of palytoxin[10]. The revolution in structure determination by NMR, however, arose only in the 1980's and beyond, with the development of multidimensional NMR. 1D H NMR shows protons at different chemical shifts based on their different chemical environments, and the relative integration shows the ratio of different types of protons. In favorable cases that information, combined with the sizes of J couplings—which represent the interactions of neighboring sites—can provided sufficient information to reconstruct a molecule from its pieces. But as the sizes of interesting molecules grow, these methods are overwhelmed by the typical overlap of peaks in the 1D spectrum. In, for example, 2D NMR, 1D NMRs are coupled together, and the cross peaks that intersects two nuclei in 2D NMRs means correlation between those two nuclei—representing, for example, the coupling between nearby ^{1}H and ^{1}H nuclei, or ^{13}C and ^{1}H. In this thesis, 1D H NMR was used to detect the amount and chemical environment of non-exchangeable hydrogens. Then, 2D NMR, including ^{1}H-^{1}H J-coupled correlation spectroscopy (COSY), or the heteronuclear ^{13}C-^{1}H J-coupled versions including single-quantum correlation spectroscopy (HSQC) and multiple-bond correlation spectroscopy (HMBC), were used for the detection of carbons and the correlations between coupled hydrogens, and the correlations of carbons and hydrogens[11]. After all of these experiments, spectra were analyzed and the structure of the unknown molecule was elucidated. We present the spectroscopy and logical thought processes involved in the analysis in what follows.

CHAPTER 2 NMR & MS EXPERIMENTS: RESULTS

2.1 1D H NMR Results

The 1D H NMR spectrum (**Fig. 1**) attached on page 4 showed 29 non-exchangeable protons. From this spectrum, some structural features were shown. The pattern shown above 7 ppm clearly indicated the existence of a mono-substituted phenyl ring where the atom of the attachment did not change much of the polarity on the ring as it integrated to 5 with little separation between each proton peaks. The peaks between 3.47 ppm and 3.78 ppm clearly showed the structure of CH_n-X, where X was nitrogen or oxygen. The peaks at 4.6 ppm and 5.83 ppm appeared in the region most often associated with C-H bonds where the C atoms participated in sp^2 hybridized bonds, or perhaps $CH\text{-}X_m$ where X was oxygen, nitrogen or other electron withdrawing groups and $m > 1$. Far upfield, at 1.2 ppm, two doublets that integrated to represent a total of 6 ^1H sites which were clearly two different methyl groups that coupled to a methyne.

Based on the assignments above, we proposed pieces that were elements of the total structure. 1D NMR was largely inadequate to provide a basis for an unambiguous set of attachments between these small building blocks. This was the motivation for the more extended multi-dimensional NMR experiments which we discuss below.

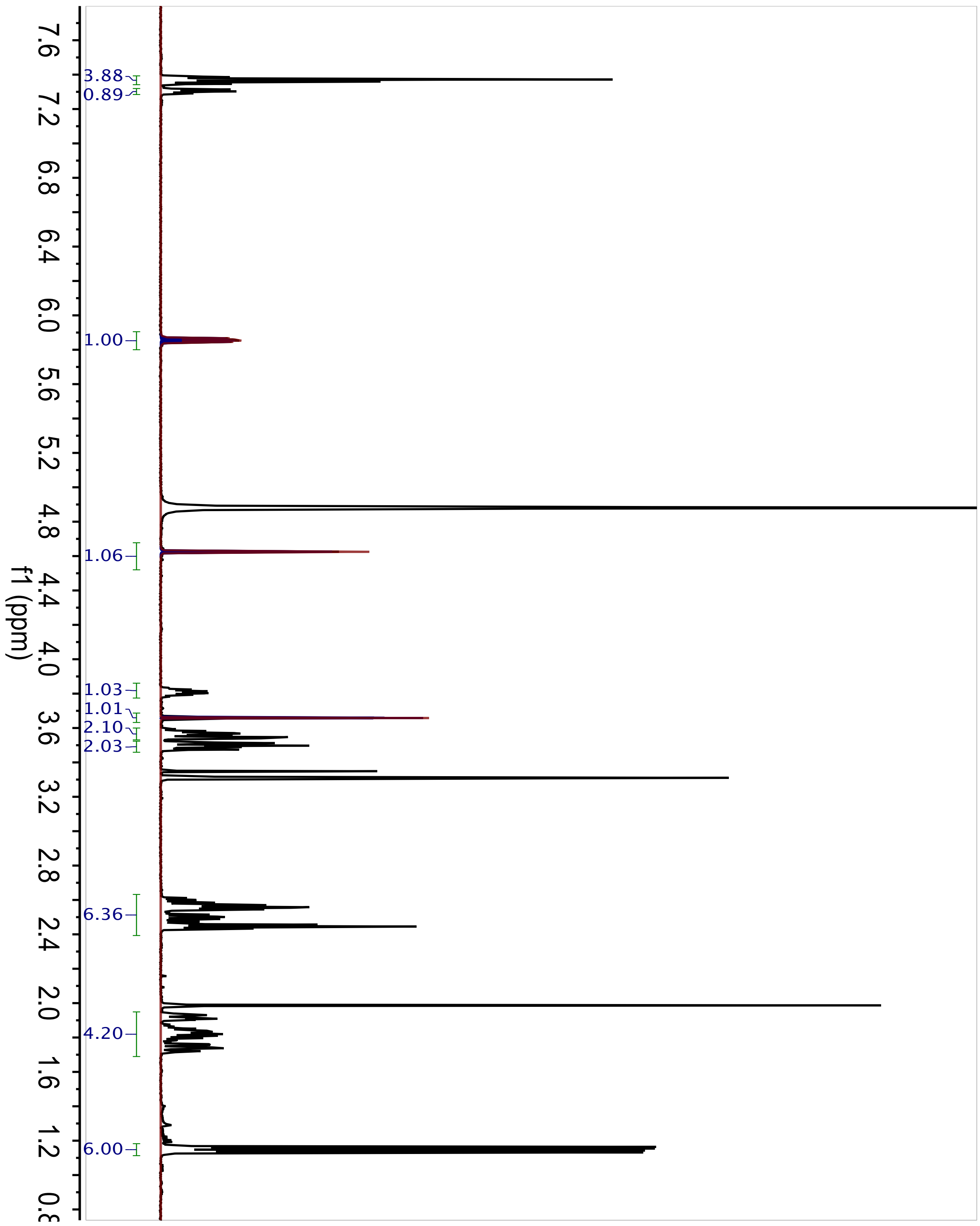
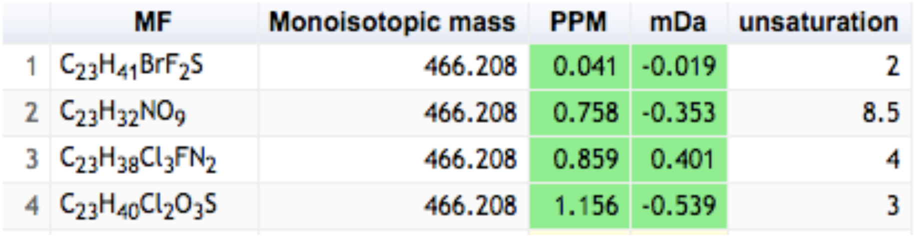


Figure 1. 1D H NMR of the unknown molecule

2.2 Results of 2D NMRs and HRMS

For 2D NMR, HMBC was taken firstly. Cross peaks in HMBC meant that the carbon and hydrogen were within the distance of 3 bonds. HMBC was designed by firstly exciting hydrogens, and then transferred the magnetism from the hydrogen to carbons that were 2-3 bonds away, so that all the carbons would show up based of the difference in the chemical shifts. In theory, a carbon that was surrounded with no hydrogen within a distance of 2-3 bonds might exist, but it was very unlikely for an organic molecule. HMBC was used to build up the carbon skeleton based on the correlations between the hydrogen on one carbon and other carbons around this carbon.

The HRMS taken in negative mode showed a molecular mass of 466.20806. The 2D HMBC attached on page 12 showed 21 carbon signals. However, knowing the existence of the phenyl ring—which if singly-substituted should have 2 pairs of equivalent ^{13}C sites—there should be 23 carbons in total. Combining that with the high resolution molecular mass determination. With only up to 3 decimal places were allowed in the website, the following molecular formulas were generated, within a 2-ppm error, from chemcalc.com:



	MF	Monoisotopic mass	PPM	mDa	unsaturation
1	$C_{23}H_{41}BrF_2S$	466.208	0.041	-0.019	2
2	$C_{23}H_{32}NO_9$	466.208	0.758	-0.353	8.5
3	$C_{23}H_{38}Cl_3FN_2$	466.208	0.859	0.401	4
4	$C_{23}H_{40}Cl_2O_3S$	466.208	1.156	-0.539	3

Figure 2. Molecular formula results from HRMS

Given the existence of a phenyl ring, the unsaturation number had to be at least 4, which ruled out formula 1 and formula 4. As no isotope pattern indicated the existence of halogen atoms, formula 3 was ruled out. Finally, the molecular formula $C_{23}H_{32}NO_9$ was chosen. The degree of unsaturation not being a whole number indicated a bond breakage, resulting only a fragment of the molecule[12]. As it was taken in negative mode, it was most likely that the fragment was the deprotonated molecule so the molecular formula should be $C_{23}H_{33}NO_9$ with a degree of unsaturation of 7 and possibly an acid moiety since deprotonation occurred. From HMBC we found three carbons in the region reserved for carbonyls, at 175.14 ppm, 173.44 ppm and 172.96 ppm. These represent either esters, amides or carboxylic acids. Combing 3 carbonyls with the phenyl ring constituted a degree of unsaturation of 7, meaning that double bond was non-existed. Given that this molecule had 29 non-exchangeable protons, there should be 4 exchangeable protons.

Then, HSQC and COSY spectra were taken. HSQC were designed to have 1-bond magnetization transfer from the excited hydrogen to the attached carbon, so that any cross peak meant direction attachment of the hydrogen and carbon. This information clarified the prior results, as some cross peaks which appeared in the HSQC also appeared in the HMBC. Using this information, we therefore determined which cross peaks represented C-H bonds, and which allowed us to clarify other nearby C atoms. Of course, quaternary carbon sites, which had no hydrogen attached, did not appear in this spectrum and must be identified via the HMBC spectrum. COSY, on the other hand, were designed to have magnetization transfer from the excited hydrogen to other coupling hydrogens. Cross peaks in COSY simply denoted coupling. Normally in a molecule, strong COSY signals meant two hydrogens that were 3 bonds away, meaning that the two attaching carbons were adjacent.

Based on the HSQC (**Fig. 5**) on page 12, COSY (**Fig. 3**) on page 10 and HMBC (**Fig. 4**) on page 11 spectra, the following table (**TABLE 1)** summarized the results:

No.	δ C (ppm)	No. of H on C	δ H (ppm)	COSY	HMBC
1	175.14	0			
2	173.44	0			

3	172.96	0			
4	138.25	0			
5,6	128.20	1	7.34	7, 8, 9	7, 8, 9
7	127.90	1	7.28	5, 6, 8, 9	5, 6, 8, 9
8,9	126.11	1	7.35	5, 6, 7	11, 5, 6, 7
10	96.01	1	4.6	14	12, 14, 17
11	74.23	1	5.83	16	3, 4, 8, 9, 16
12	70.19	1	3.78	22, 18	10, 20
13	69.92	1	3.54	23	15
14	68.37	1	3.64	17, 10	15
15	66.92	1	3.47	17	11, 13
16	44.19	2	3.52, 3.47	11	2, 4
17	34.42	2	1.90, 1.71	14, 15	
18	31.89	2	1.82, 1.71	12, 20	
19	30.15	2	2.42	21	1, 2, 21
20	30.07	2	2.58, 2.42	18	3, 12, 18
21	28.95	2	2.53	19	1, 2, 19
22	17.71	3	1.11	12	12, 18
23	16.65	3	1.14	13	13, 15

TABLE 1. Summary of NMR results

Given the chemical shift assignments above, C_1 to C_3 would be the three carbonyls, and C_4 to C_9 would be the phenyl ring. C_{16} with a chemical shift of

44.19 ppm represents a CH_2-N group. C_{11} to C_{15} with chemical shifts within the range of 65 ppm and 75 ppm would be C-O carbons. C_{10} with a chemical shift of 96.01 represents a carbon that has two attached heteroatoms, as well as an attached H atom. C_{17} to C_{21} would be aliphatic methylene carbons. C_{22} and C_{23} would be two methyl groups. They methyl groups had different chemical shifts both for ^{1}H and ^{13}C because the local chemical environments were different as they had different sourrounding atoms.

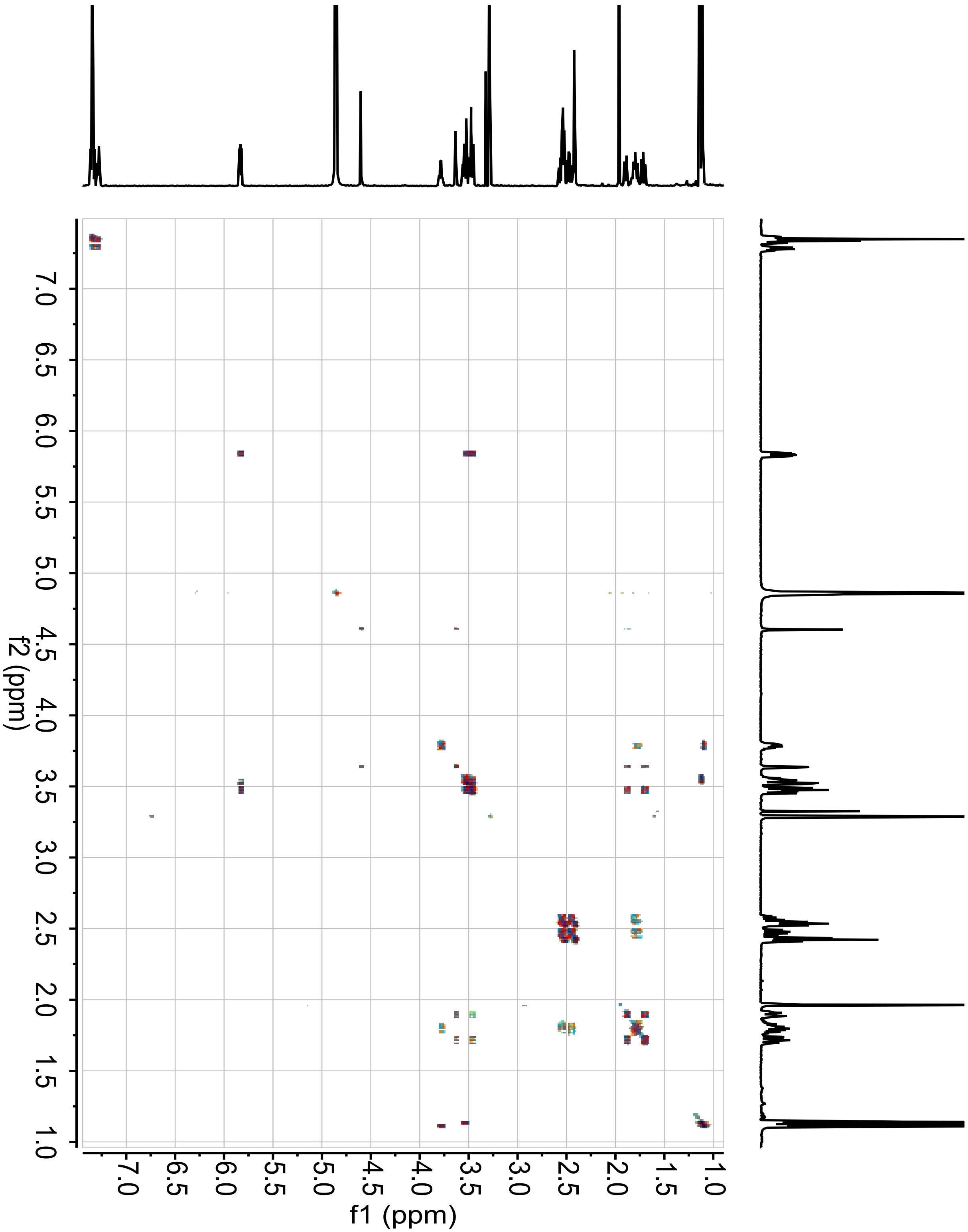


Figure 3. 2D COSY NMR of the unknown molecule

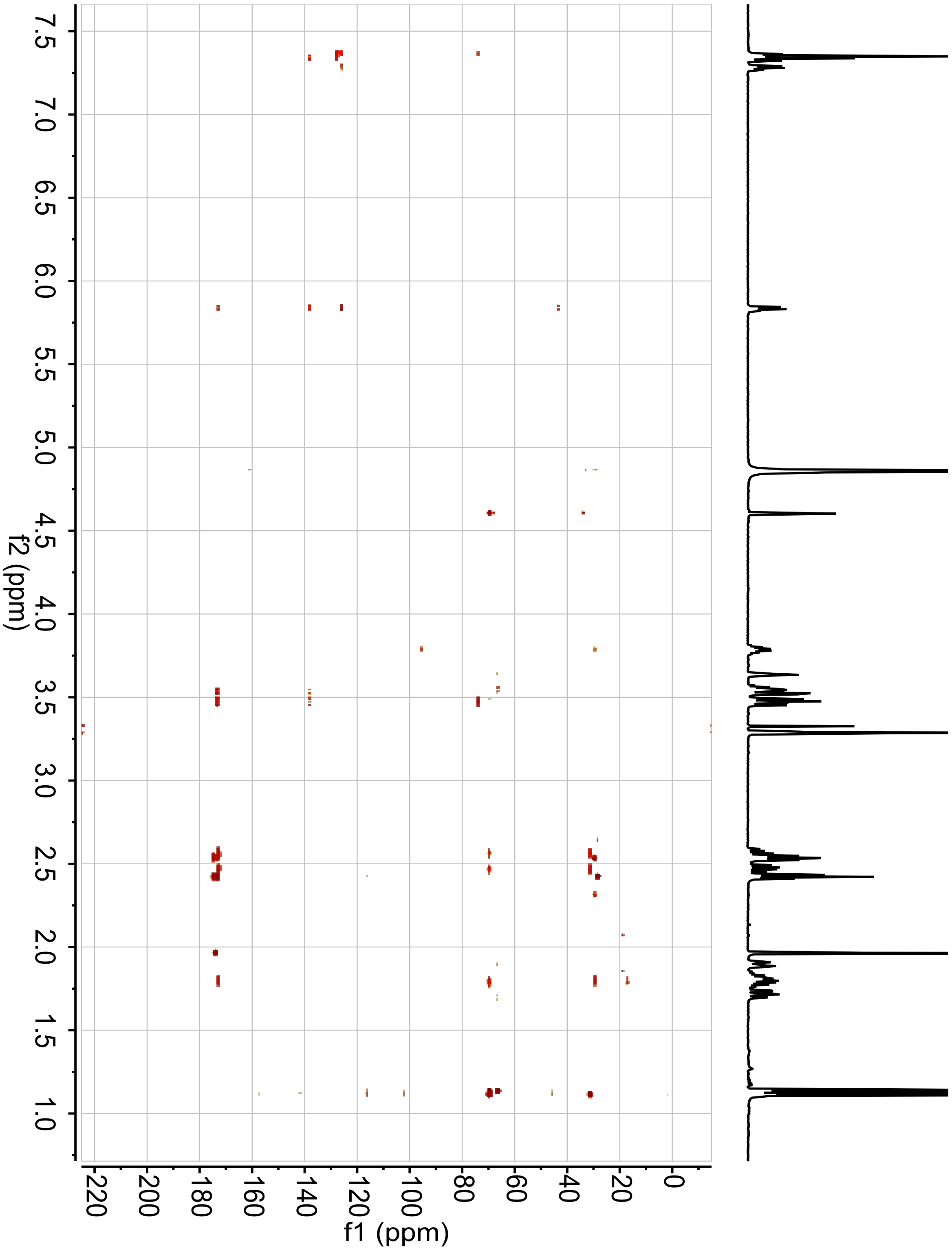


Figure 4. 2D HMBC NMR of the unknown molecule

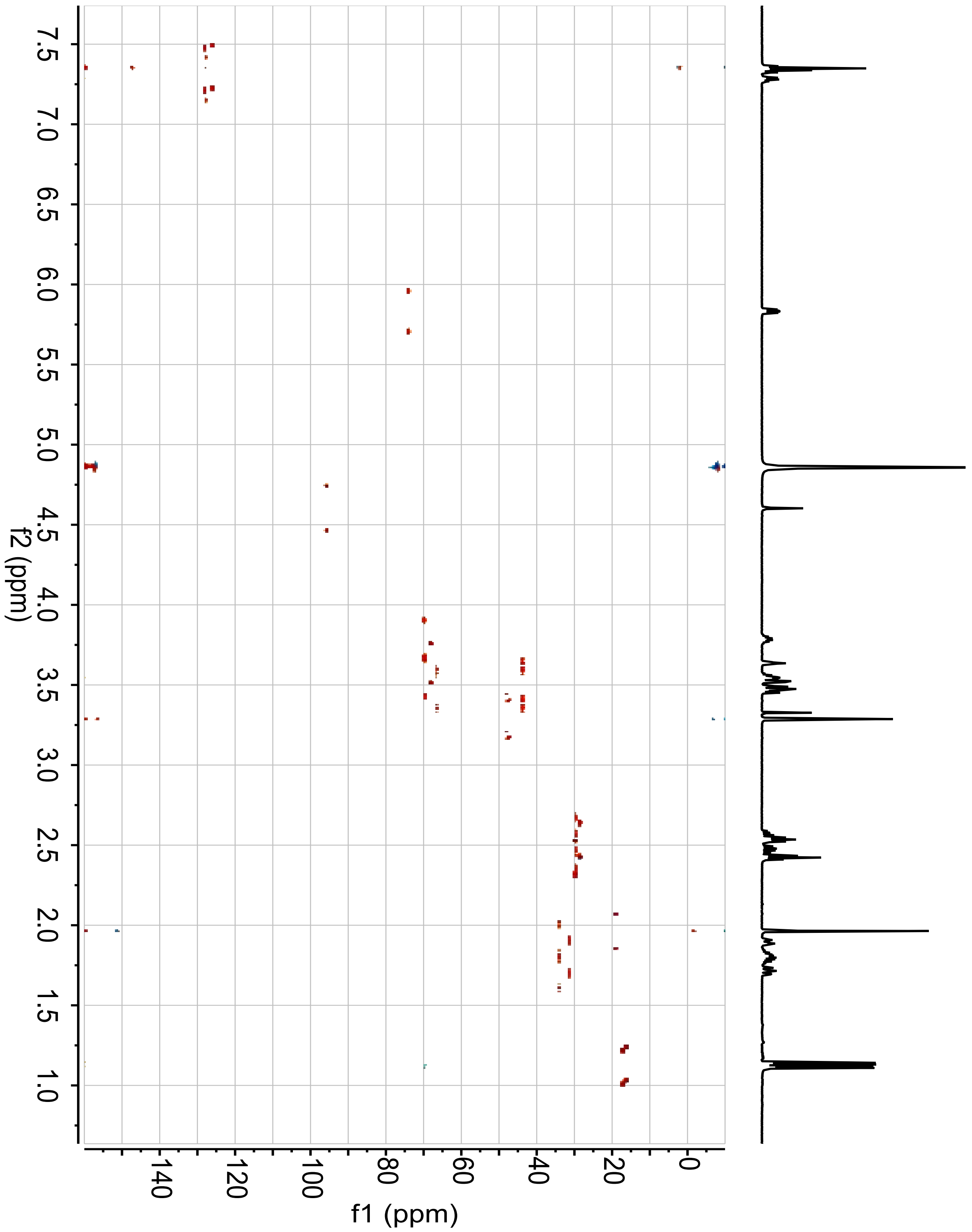


Figure 5. 2D HSQC NMR of the unknown molecule

CHAPTER 3 Elucidation of a Proposed Structure of the Unknown Molecule

Our task in this chapter is to reconstruct a molecule based on all the pieces of information discussed in the prior two chapters. The structure could be initiated from literally anywhere though it makes sense to work from a group which was "terminal" in the molecule, i.e. a group which was connected to the rest of the molecule by only a single bond. This supported starting from a methyl group; we would begin with the methyl group identified as C_{22} above. H_{22} had COSY signal to H_{12}, H_{12} had COSY signals to H_{18} and H_{22}, and HMBC signal to C_{10}, H_{18} had COSY signals to H_{20}, H_{20} had HMBC signal to C_3, and H_{11} had HMBC signals to C_3. The following partial structure was thus concluded:

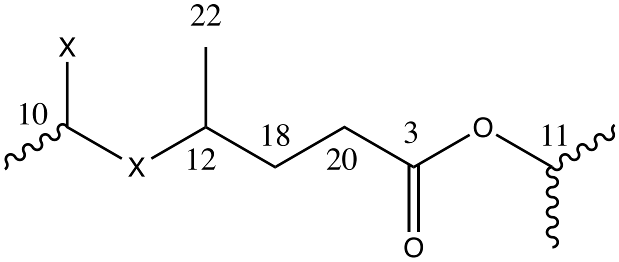


Figure 6 Partial Structure 1

Then, starting from C_{11} and going right, C_{11} had HMBC to C_8, $C_{9,}$ C_3 and C_4, H_{11} had COSY signal to H_{16}, H_{16} and H_{19} had HMBC signal to C_2, the following structure was concluded:

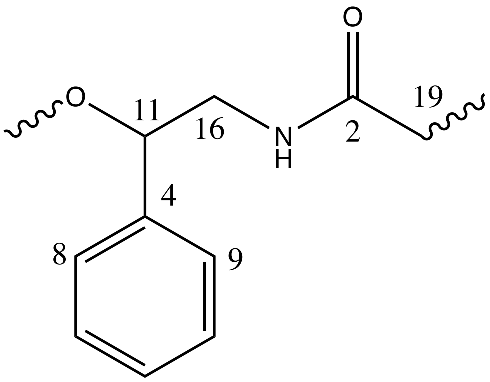


Figure 7 Partial Structure 2

Next, starting from C_{19}, H_{19} had COSY signal to H_{21}, H_{21} had HMBC to C_1, the following structure was concluded:

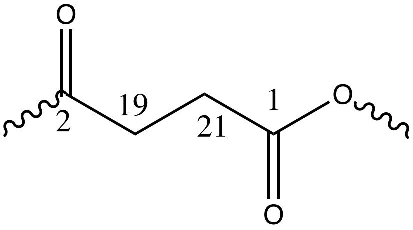


Figure 8 Partial Structure 3

Finally, back to **Fig. 6** and starting from C_{10} and going left, since the nitrogen was taken, the two heteroatoms had to be two oxygens. Given that H_{10} had COSY signal to H_{14} with a small coupling constant, H_{14} had COSY signals to H_{10} and H_{17} with three small coupling constants, H_{17} had COSY signals to H_{15} and from COSY, H_{15} had two large coupling constants and one small coupling constant. Coupling constants were dependent on the dihedral torsion angles between two protons[13]. Large coupling constants indicated a large dihedral angel, which in this case, meant that two protons were in a di-axial position and small coupling constant indicated at least one proton was in an equatorial position[14]. H_{23} had COSY signal to H_{13}, H_{10} had HMBC signal to C_{13} and H_{23} had HMBC signal to C_{15}, the following structure was proposed:

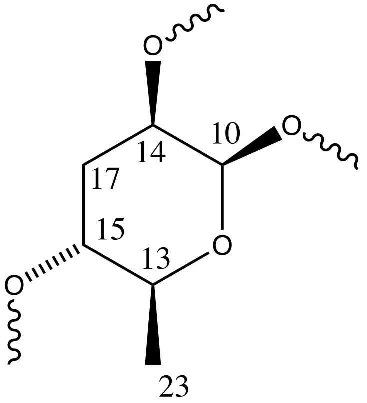


Figure 9 Partial Structure 4

Partial structures 1 to 4 contained all 23 carbons. Combining partial structure 1 to 4 and the fact that there were 4 exchangeable protons, the final structure was concluded:

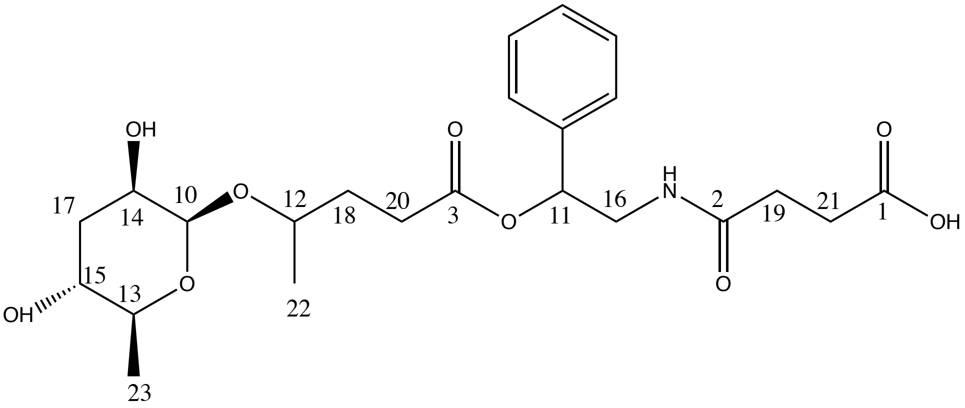


Figure 10 Proposed Structure

CHAPTER 4 DISCUSSION & CONCLUSION

In this thesis, the NMR spectra and HRMS of an unknown molecule were analyzed and a structure was proposed based on the analysis.

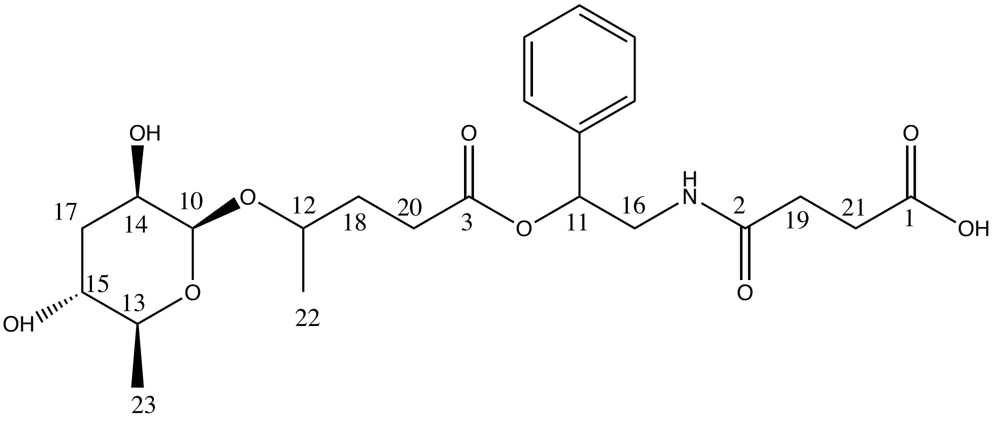


Figure 10 Proposed Structure

The structure overall was well-supported. However, the determination of absolute stereochemistry could be achieved if nuclear Overhauser effect spectroscopy (NOESY), whose cross peaks meant spatial proximity, was provided. For now, only the relative stereochemistry on the nonaromatic 6-membered ring was solved, but the configuration of the two stereocenter on the carbon chain remained unsolved. With NOESY provided, cross peaks were expected to see among not only hydrogens on the ring, indicating they were on the same phase of the ring or not, but also along the carbon chain so that any diastereotopic hydrogens would be clearly shown, and by looking at NOESY signals, i.e. the spatial proximity of two hydrogens, the absolute stereochemistry could be solved. For example, if H_{12} had NOESY signal to H_{10}, H_{12} and H_{10} would be on the same phase, indicating that H_{10} was pointing into the page, or if H_{22} had NOESY signal to H_{10}, H_{12} would be pointing out of the page since H_{22} would be pointing into the page. But currently, only the relative stereochemistry on the ring was resolved, and further experiments were needed if the absolute stereochemistry needed to be solved.

BIBLIOGRAPHY:

1. Moore, R.E. & Scheuer, P.J. Palytoxin: a new marine toxin from a coelenterate. *Science* 1971, 172, 3982, 495-498.
2. Downard, K. Mass Spectrometry – A Foundation Course. *Royal society of Chemistry* 2004, doi: 10.1039/9781847551306.
3. John D. Roberts. Nuclear Magnetic Resonance: applications to organic chemistr*y*. McGraw-Hill Book Company, 1959. ISBN:9781258811662.
4. Mass spectrometry. (n.d.). In *Wikipedia* Retrieved July 16, 2018, from https://en.wikipedia.org/wiki/Mass_spectrometry
5. Ho, CS; Chan, MHM.; Cheung, RCK; Law, LK; Lit, LCW; Ng, KF; Suen, MWM & Tai, HL. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin. Biochem. Rev.* 2003, 24, 1, 3-12.
6. Naomi, L.S. Introducing Graduate Students to High-Resolution Mass Spectrometry (HRMS) Using a Hands-On Approach. *J. Chem. Educ.* 94, 12, 1978-1982.
7. Xian, F.; Hendrickson, C.L. & arshall, A.G. High Resolution Mass Spectrometry. *Anal. Chem.* 2012, 84, 708-719.
8. Paudler, W. Nuclear Magnetic Resonance. *Boston: Allyn and Bacon Chemistry Series* (1974), 9-11.

9. Pople, J.A.; Bernstein, H.J. & Schneider, W.G. The Analysis of Nuclear Magnetic Resonance Spectra. *Can. J. Chem.* 1957, 35, 65-81.
10. Ko, S.S.; Klein, L.L.; Pfaff, L.P. & Kishi, Y. Synthetic Studies of Palytoxin. Stereocontrolled, Practical Synthesis of the C.101 – C.115 Segment. *Tetrahedron Lett.* 82, 23, 43, 4415-4418.
11. Aue, W.P.; Bartholdi, E.; Ernst, R.R. Two-dimensional Spectroscopu. Application to Nuclear Magnetic Resonance. *J. of Chem. Phys.* 1976, 2229 - 2246.
12. Badertscher, M.; Bischofberger, K.; Munk, M.E. & Pretsch, E. A Novel Formalism to Characterize the Degree of Unsaturation of Organic Molecules. *J. Chem. Inf. Model.* 2001, 41, 4, 889
13. Karplus, M. Vicinal Proton Coupling in Nuclear agenetic Resonance. *J. Am Chem. Soc.* 1963, 85, 18, 2870- 2871
14. Minch, M.J. Orientational Dependence of Vicinal Proton-Proton NMR Coupling stants: The Karplus Relationship. *Concepts in Magnetic Resonance.* 6, 41-56. Doi: 10.1002/cmr.1820060104