Isolation and Structural Characterization of Defensive Natural Products from *Sinularia* Soft Coral

by

Heather Rainbow

Submitted in partial fulfillment of the requirements for graduation as an Honors Scholar at Point Loma Nazarene University, San Diego, California, May 4, 2022.

Approved by _____Katherine Maloney_____
[Mentor Name]

_____Matthieu Rouffet_____
[Committee Member Name]

_____Greg Crow_____
[Committee Member Name]

Date_____5/07/2022_____

Table of Contents

Abstract

Octocorals such as those belonging to the genus *Sinularia* commonly employ two methods of defense against predation. The first is categorized as physical defense, and consists of calcium carbonate spines called sclerites which are very sharp and hard for predatory fish to bite. The second category is chemical defense, which consists of small molecules called natural products synthesized by the coral itself which incur antifeedant or cytotoxic effects on predators. In this project, we sought to apply statistical methods to concentrations of natural products within soft coral samples, in order to suggest possible novel chemical defense compounds. By comparing our measure of physical defendedness (namely, percent mass of sclerites) with the peak areas of organic compounds within the coral samples, we identified compounds which demonstrate an inverse correlation with physical defendedness, as we might expect for a chemical defense compound. Three top "suspect compounds" were identified using PatternSearch in MetaboAnalyst, identified as features 3701, 2051, and 3705. Additionally, molecular networking in GNPS resulted in the discovery that many of the inversely correlated compounds are structurally similar to each other, and largely found in Clade 4 of the *Sinularia* genus, the least physically defended clade. However, in order to assess the significance and applicability of these correlations, further work is needed. We plan to isolate and acquire cytotoxicity data on suspect compounds in the future.

Introduction

Certain marine species provide a unique glimpse into the world of defense against predation. Octocorals employ a unique method of structural defense by growing sharp calcium carbonate spines called sclerites. The growth of these sclerites are so unique to each species of coral that taxonomists use their characteristics to describe and distinguish species[2]. However, aside from just differentiating species, these sharp spines function as a physical method of defense, as they prevent predatory fish from chomping down on the[7] soft tissue of the coral's body.

Additionally, soft corals such as *Sinularia* can synthesize chemical defense natural products, which may serve multiple purposes for the organism but also induce cytotoxic effects

on predators. Many of these secondary metabolites belong to the family of isoprene[5] derivatives known as terpenoids. One example of these is “pukalide,” named for its emetic effect on fish. These natural products not only deter the aforementioned predators but also help the corals survive algal blooms and microbial growth. Thus, some of these natural products demonstrate antimicrobial or anti- inflammatory properties as well as cytotoxicity against predators.

Because all organisms fundamentally need to conserve energy, a trade-off theoretically might exist between types of defense measures. In other words, a coral that is highly physically defended might not waste energy synthesizing chemical defense natural products. Conversely, a very soft and spongy coral, with almost no sclerites, would possibly require a different defense strategy and therefore synthesize higher levels of cytotoxic compounds.

Figure 1 below shows the phylogenetic tree of the octocoral genus *Sinularia* and the ancestral relationships between clades. Clade 1, containing the very sclerite-rich species *S. brassica,* tends to be highly physically defended, while clade 4 contains *S. flexibilis,* a very soft and squishy species which has been well studied as a rich source of chemical defense compounds.

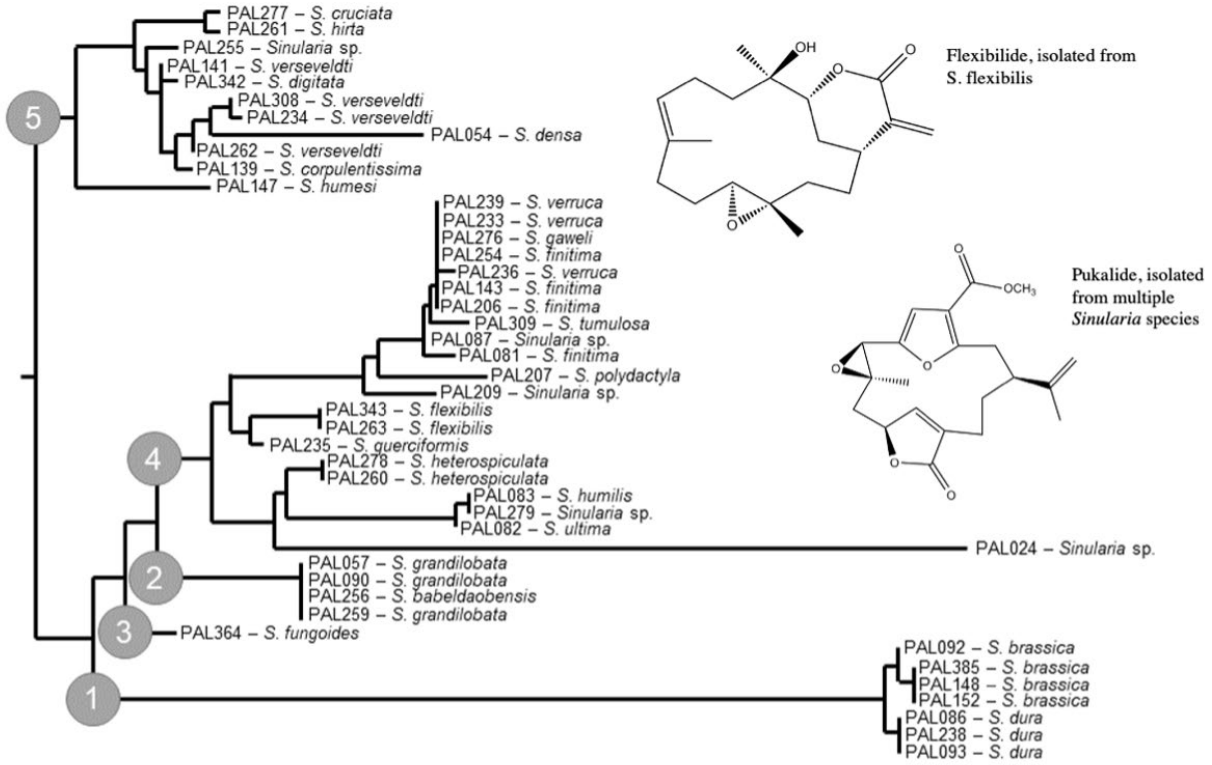


Figure 1: Phylogenetic tree showing ancestral relationships between species within the genus Sinularia. Clades, which are small closely related groups of species, are numbered 1-5.

This suggested trade-off in energy leading to a hypothesized inverse correlation has been explored previously. *Sammarco et al* investigated in 1992 an order of octocorals, and suggested that a correlation between physical defense and cytotoxicity to fish might exist; however, they recommended interrogating this at a higher degree of taxonomic resolution (i.e. at the genus or family level). Therefore, this project zeroes in on one genus within that order: *Sinularia.* Additionally, *Santacruz et al* investigated in 2019 the relationship within soft coral between metabolomic fingerprints and cytotoxicity; however, rather than using statistics to pick out "suspect" compounds and then confirming whether that suggestion was correct, these researchers evaluated the cytotoxic effects of all of their samples before commencing with statistical tests on their LCMS data such as Principal Component Analysis (PCA) and Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA). They found that their data analyses were valuable in highlighting and identifying the cytotoxic metabolites, and an OPLS-DA plot

demonstrated separation between compounds of higher and lower cytotoxicity according to the cell line assays. This project seeks to develop knowledge of this separation by first trying to predict suspect features using data analysis, and later work will investigate the cytotoxicity of any isolated compounds. In 2017, *Farag et al* sought to investigate *Sinularia* soft coral extract NMR fingerprint-determined diterpene levels and used statistical tools to determine if natural separation occurred between compounds with high diterpene levels and lower diterpene levels (chemical defense natural products are, as mentioned previously, frequently terpenoids). They additionally conducted assays to determine cytotoxicity IC50 values for each extract, and they did not find a conclusive correlation between cytotoxicity and diterpene levels. The authors suggested that perhaps NMR metabolomics cannot easily predict this correlation.

The explosion of the field of metabolomics in recent years has inspired the development of new computational tools for managing metabolomic data (which is inherently *much* more complex than genomic data). Much of the metabolomics research relies on mass spectrometry (MS) - an especially sensitive and specific technique for identifying and quantifying compounds in complex mixtures. This explosion of mass spectrometry data has in turn prompted the development of new computational tools for analyzing and comparing complex MS data across large numbers of samples. MZMine is one such tool. This open-source LC-MS processing software is used in metabolomics research to identify recurring compounds within large LCMS data sets - a process known as feature finding. LC-MS data acquisition followed by analysis in MZMine provides another option for investigating the proposed correlation between physical and chemical defense in *Sinularia*. MZMine computationally performs peak detection, alignment, and deconvolution, allowing the user to compare many chromatograms to each other and pulling out recurring retention times and masses, along with the associated peak areas.

We hypothesize that chemical defense compounds within coral extracts inversely correlate with physical defendedness in a collection of *Sinularia* specimens collected in Palau. To test this, we performed statistical analyses on peak areas of compounds found in the coral samples, identified using MZMine, to compare the abundance of certain molecules with a measure of physical defense (percent mass of the total mass of coral). Compounds identified as being inversely correlated with percent mass of sclerites will be targeted for isolation, structure elucidation using NMR, and ecological testing.

Experimental

Sample Acquisition

50 samples of *Sinularia* soft coral were collected from the Republic of Palau using SCUBA in late May 2010 and frozen at -20 degrees C. Each specimen was transported on dry ice to Claremont, California and then to San Diego, California in December 2011. These samples were stored in the lab at -20 C at atmospheric pressure until LC-MS/MS analysis nine years later.

Sample Preparation: LCMS/MS Data

100mg of each specimen was lyophilized, ground in a ball mill homogenizer and sonicated in 1:1 DCM:MeOH for 15 minutes, and centrifuged for 10 minutes at 14000 rpm. The supernatant was transferred to replicate 96-well plates and the solvent removed by rotary evaporation. The residue was resuspended in 200 uL of MeOH containing 2 uM sulfamethazine as internal standard, sonicated to 10 minutes, and centrifuged for 10 minutes at 2000 rpm. The supernatant was diluted 2-fold in MeOH containing 2 uM sulfamethazine a total of four times to give a final concentration of .0625x relative to the original resuspended extract. 2uL of this diluted extract was injected for LC-MS/MS analysis using a Kinetex C18 column (1.7 μm × 50 mm × 2.1 mm, Phenomenex), coupled to a Maxis Impact HD Q-TOF mass spectrometer (Bruker Daltonics). The column was equilibrated with 25% solvent B (LC-MS grade MeCN, 0.1% formic acid) for 1 min, followed by a linear gradient from 25% B to 100% B in 8 min, held at 100% B for 2 min, then adjusted back to 25% B over 0.5 min and maintained at 25% B for 1 min. The flow rate was maintained at 0.5 mL/min throughout the run. MS spectra were acquired in positive ion mode in the range of 50−2000 m/z. Data were collected using a data dependent acquisition method with an advanced stepping method and assessed for quality.[3] Raw data files were converted to mzXML format using Bruker Compass DataAnalysis 4.2 software.

Sample Preparation: Percent Mass Data

In order to quantify a measure of physical defendedness, the remaining parts of the coral samples were weighed in 2011 to obtain an initial mass. After extraction of organic material similar to the procedure previously described (repeated sonication and decanting using 1:1 DCM:MeOH as solvents), the remaining inorganic fragments (the sclerites) were soaked in bleach for 3 days to ensure dissolving of any remaining organic material. They were then washed in acetone and oven-dried to obtain clean, dry sclerites. These sclerites were massed, and their length and overall size analyzed using ImageJ. The mass of the sclerites by themselves was divided by the initial mass of the sample to calculate a percent mass of sclerites.

Data Analysis: Feature-Based Molecular Networking

Feature detection, grouping, and alignment were performed using MZmine2, following the feature-based molecular networking documentation.[4] 4,453 features were identified. Detailed MZmine2 parameters are described in the Appendix (*Table A1)*. The CSV file (quantitative feature table) and MGF file (MS/MS spectra for each feature) generated in MZmine2 were uploaded and used in the feature-based molecular networking workflow in GNPS (http://gnps.ucsd.edu). The precursor ion mass tolerance and product ion mass tolerance were both set to 0.05 Da. Molecular networks were generated using 12 minimum matched peaks and a cosine score of 0.8. The molecular networking job on GNPS can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b6a4a916f5b7487da85aa30f95bc5d66. Data was visualized in Cytoscape.

Data Analysis: Feature Table in Excel

The CSV file (quantitative feature table) from MZMine 2 was modified to include percent mass data of each sample. Raw correlations were run in Excel between peak areas of identified features and percent masses using the [=correl] function. A number of interesting inverse correlation coefficients were identified, but further analysis directed by graphs such as *Figure 2* determined that the correlations were heavily influenced by outliers or only 1 or 2 data points.

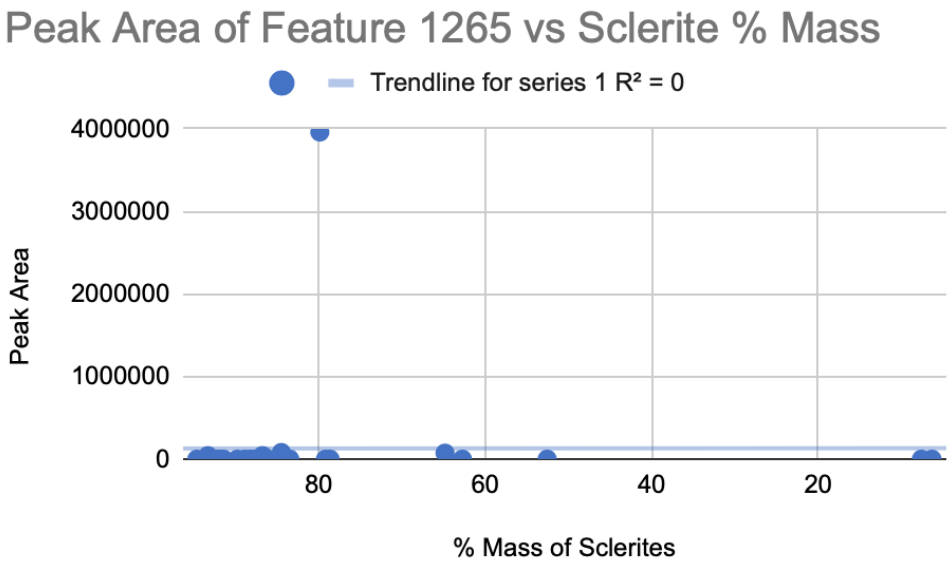


Figure 2: Graph of Feature 1265 Percent Mass of Sclerites versus feature peak area. The compound, despite having a large inverse correlation coefficient according to Excel, clearly does not consistently exhibit higher abundance when sclerite amount is low, and does not slope downward/decrease in concentration gradually when sclerite amount is high.

Additionally, some features only appear in 2 or 3 samples, leading to a sparse data matrix which might deceptively make those compounds appear to correlate with sclerite percent mass.

Data Analysis: MetaboAnalyst

To further explore possible correlations, the CSV file (quantitative feature table) from MZmine2 was modified to include both clade identity and percent mass data (organized into 4 bins of Low, Medium, High, and Very High) according to the histogram in the Appendix (*figure A1)*. The table was uploaded to Metaboanalyst (www.metaboanalyst.ca). No missing value estimation or data filtering were performed; the samples were normalized using log transform. PCA, PCoA, Random Forest, and PatternHunter correlation analysis were performed in MetaboAnalyst.

Sample Preparation: Second Round of Extraction, Isolating + Characterizing Suspect Features

Three members of clade 4 (identified as PAL343, PAL263 (both *S. flexibilis)* and PAL235 (*S. querciformis)* were selected for bulk extraction. The remaining lyophilized tissue (after prior sampling) was pooled and crushed using a mortar and pestle. Organic material was extracted using 150 mL of 1:1 DCM:MeOH in 30 mL increments. The fragments were sonicated and extracted to exhaustion. The extracts were filtered to remove particulate matter and the solvent was evaporated using a rotary evaporator to give 1.244 g of crude extract.

The extract was loaded using the dry-load method onto a silica gel CombiFlash Rf (Teledyne ISCO) flash column (4 g, x2 stacked) and eluted using a gradient from 100% hexanes to 100% EtOAc over 20 minutes, then immediately back dropped down to 0% EtOAc, then a slight gradient of 100%-80% hexanes:EtOac over 5 minutes. Similar fractions were combined as directed by TLC to give 13 fractions A-M (A: 18.9 mg, B: 127.0 mg, C: 47.5 mg, D: 27.1 mg, E: 24.3 mg, F: 59.8 mg, G: 23.4 mg, H: 49.7 mg, I: 192 mg, J: 59.4 mg, K: 26.7 mg, L: 5.2 mg, M: 35.3 mg). ^{1}H, ^{13}C, and 2D NMR spectra including HSQC and COSY of all fractions A-M was acquired on a JEOL 400 NMR spectrometer. HSQC data was used to generate a table to be uploaded to SmartNMR (See *Table A2* in the Appendix*)* to characterize fractions. According to the SmartNMR search, fraction G contained cembranoids, and fractions A and C may contain cembranoids.

Another round of LCMS data was collected on all fractions A-M. The most abundant mass in fraction G was 377.2339, which is consistent with a cembranoid structure. The candidate structure for fraction G, determined by the smartNMR search and the most abundant mass found in the LCMS, is shown below and is an estimation of the actual structure subject to further confirmation.

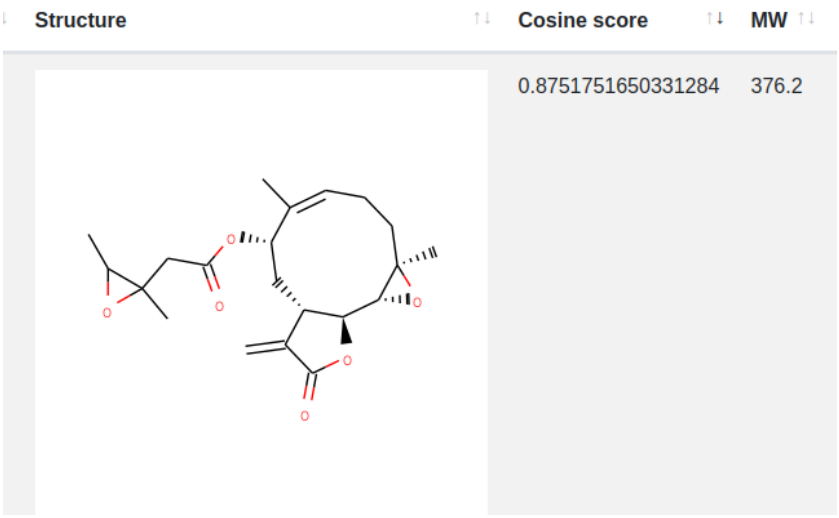


Figure 3: Candidate structure of cembranoid found in Fraction G, likely an isomer to the true structure and subject to confirmation.

For clarity, a summary of the workflow for this project can be found in the Appendix in *Figure A2*.

Results and Discussion

Data Analysis

MetaboAnalyst tests such as Random Forest, PCA and PLSDA identified "VIP features," that is, features whose levels are most strongly predictive of differences in a specified variable - in our case, abundance of sclerites. However, the features identified by these tests did not exhibit the desired inverse correlation with percent mass of sclerites. For example, *figure X* below shows a VIP feature according to the Random Forest tests; in this case, we see that the feature shows a *positive* correlation with % sclerites, rather than the sought inverse correlation.

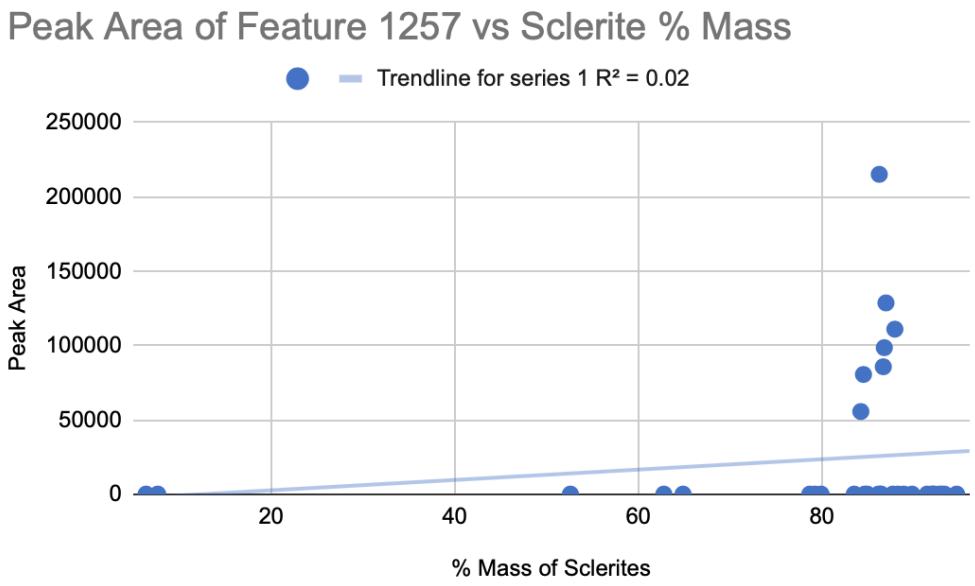
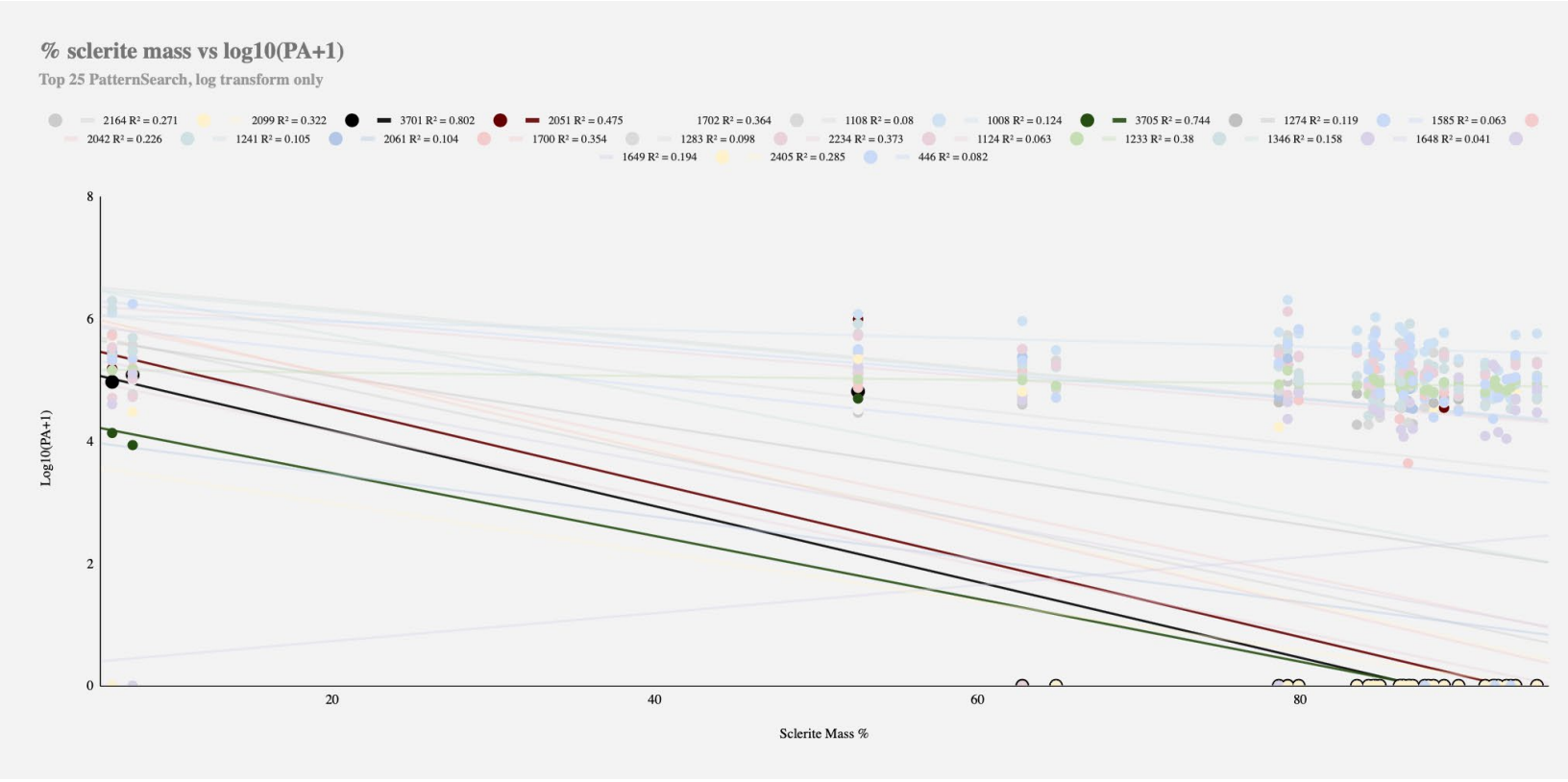


Figure 4: Feature 1257, a VIP according to RandomForest, actually increases with the amount of sclerites rather than decreases (note: this data has not been normalized). Some features identified by RandomForest followed this pattern, while others showed no correlation in either direction at all, and still others demonstrated a very slight inverse correlation. Therefore, RandomForest did not seem to consistently identify the patterns we were seeking.

One statistical test did pick out features whose correlations matched what we were looking for, and that was PatternSearch, a template matching method. The template for the computer to follow was 1-2-3-4, looking for a compound that increased in abundance from 1 (Very high percent mass sclerites, VH) to 4 (low percent mass sclerites, L). Below are the top 25 features according to PatternSearch, graphed as the amount of sclerites versus peak area.

Figure 5: Top 25 PatternSearch VIPs after log transform. Nearly all these compounds exhibit the desired slope from high abundance to low abundance as sclerite amount increases. The 3



darkened lines represent the 3 features with the highest R^2 values (features 3701, 2051, and 3705). These 3 features became our "suspect compounds," which can be targeted for isolation.

Feature Number	Retention Time	Parent Mass
2051	0.2838	212.0926
3701	6.5031	989.7314
3705	3.2684	759.3921

Table 1: Suspect compounds from the above graph which can be targeted for isolation, described by their feature number from MZMine, retention time, and parent mass.

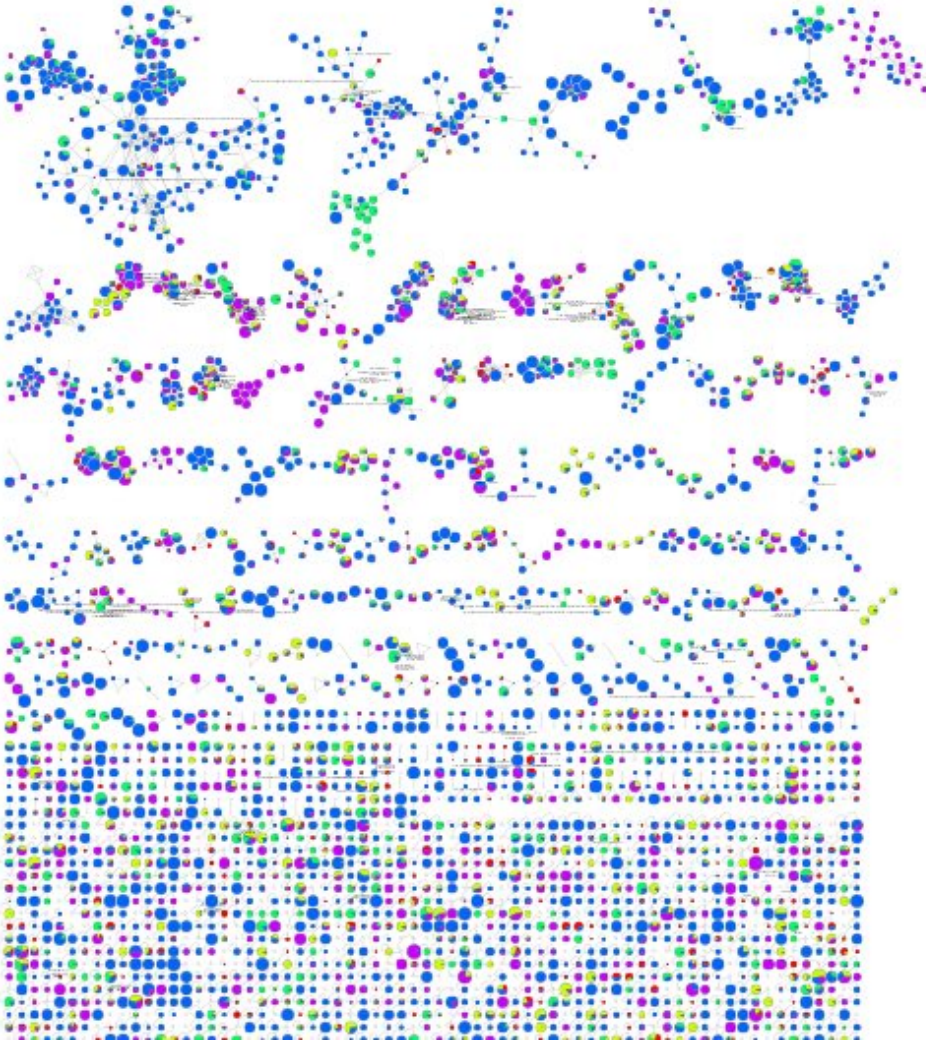
This graph in *Figure 5* is promising, and is certainly the best we found out of all the statistical methods used due to its apparent decreasing slope and relatively larger R^2 values associated with each correlating feature; however, it is limited in that the abundances of sclerites are not gradual but rather clustered around 0-5%, 55%, 65%, and >80%. Ideal data would include more points across more variance in sclerite abundance. Additionally, the darkened

points in the top middle and right of the graph are obvious outliers, and we do not observe many data points in the lower peak area range that are non-zero. Therefore, the correlation is observed but not as strongly as we had predicted.

Molecular Networking in GNPS

To explore possible structures, the LCMS data set (which also contains MS/MS data) was subjected to molecular networking in GNPS. LCMS data describes retention times and masses of compounds, while MS/MS data describes fragmentation patterns of the molecules in the samples. Similar molecules fragment in similar patterns, and molecular networking compares those fragmentation patterns to each other as well as to library compounds to look for similarities to each other or matches to the database. The molecular network quantifies the amount of similarity between various features (compounds) identified by MZMine and then connects similar-fragmenting compounds into clusters of presumably structurally related molecules. The molecular network for the feature table, containing a total of 4453 unique features, is shown below in *Figure 6.*

Figure 6: Feature-based Molecular Network[8], visualized in Cytoscape 3.8.2. Some self-looped features are omitted from the bottom of the figure to improve visualization. The largest cluster on the top left is shown zoomed-in below in Figure X. Network nodes are sized according to the PatternSearch coefficient, the larger nodes being those features which exhibited the strongest inverse correlation with abundance of sclerites. Nodes contain pie charts which communicate



which clades that feature was found in, according to the following key: Red = 1, Yellow = 2, Green = 3, Blue = 4, Purple = 5.

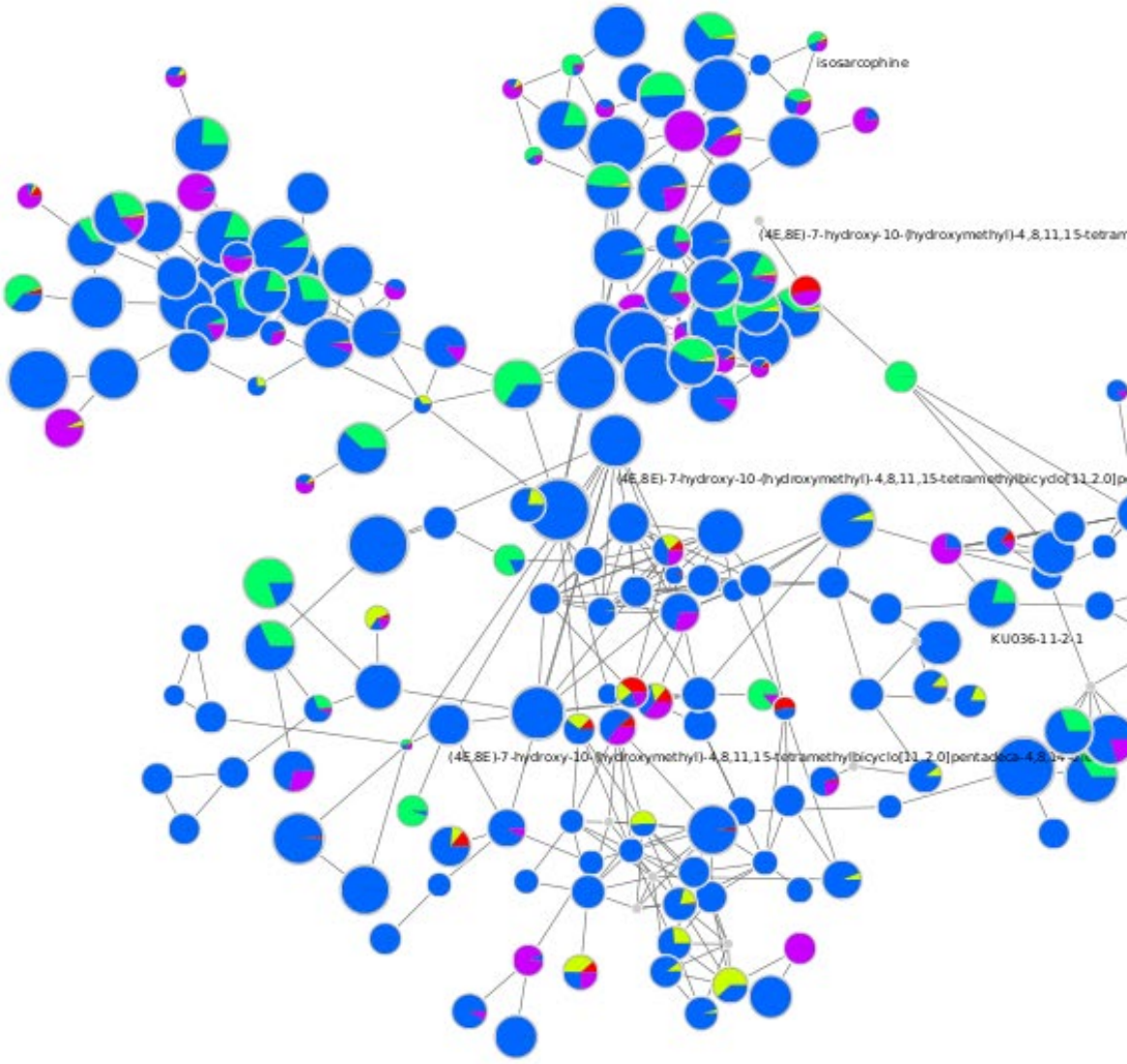


Figure 7: Zoomed in view of the largest network cluster. 4 nodes matched to library cembranoid structures, one of which is shown below, suggesting that the rest of the cluster is cembranoid-like and possibly relatives of the library structures.

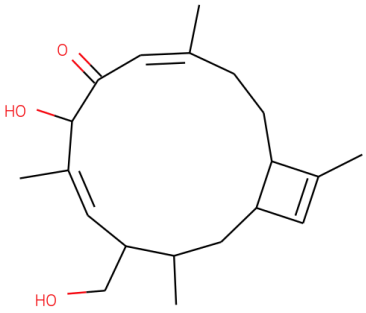


Figure 8: Structure of library compound which networked to 3 nodes in the cembranoid cluster.

Interesting facets of the network include the partial clustering of highly inversely correlating compounds together, meaning that those compounds are structurally similar (and are likely cembranoids, which we know from the literature is frequently the structural family of chemical defense compounds). Additionally, most of the larger nodes are predominantly blue, indicating that they are mostly found in clade 4. This is consistent with our hypothesis, since clade 4 is the least physically defended clade. We can also recognize from the network that clade 4 is especially chemically prolific - the network as a whole contains an abundance of blue or largely blue pie charts. Many clusters of molecules are somewhat clade-specific– we observe sections of the network that are mostly green, purple, or yellow. However, some individual molecules are, fascinatingly, found in all clades, demonstrated by the pie charts which contain all colors in the key (red, purple, green, blue and yellow).

Conclusions and Future Work

Although the hypothesized correlation is not observed very definitively in this data, that does not necessarily mean the energy trade-off idea is invalid. One possible explanation could be that cytotoxic natural products are only found in the species that require them, and do not gradually decrease in concentration as abundance of sclerites increases, but rather simply are not synthesized by those highly physically defended corals at all. Additionally, it is possible that the percent mass of sclerites is not a fit way to measure physical defendendess, and the correlation might be better observed by comparing the sharpness or length of the sclerites to the peak areas of the identified features.

Another possible method to look for the correlation could be to zero in on one species, *S. flexibilis,* and acquire a variety of samples of that one species. One species may exhibit a slight variance in its amount of sclerites across individual corals, and this would allow us to look for a correlation between feature peak areas and abundance of sclerites at the highest possible phylogenetic resolution.

The next step in this project is to further purify the fractions that contain cembranoids using HPLC. Ideally these fractions will also be subjected to another round of LCMS acquisition in the same conditions and column on the original instrument, the Maxis Impact, to allow comparison of isolated compounds' retention times to suspect features' retention times. The MS/MS spectra can also be used to match to the suspect cembranoids from this analysis. The structures of isolated cembranoids will be elucidated and compared to known chemical defense compounds. Additionally, isolated compounds will be sent to Dr. Valerie Paul at the Smithsonian Marine Station. Dr. Paul is a chemical ecologist who has agreed to assay our purified cembranoids for antifeedant activity against fish and other coral predators. Results of interest would include finding that a known chemical defense compound does exhibit the hypothesized correlation, or that the suspect features do turn out to be chemical defense compounds when tested against coral predators. Either such result could serve as an illustration of the potential value of statistical analyses in helping to understand chemical ecological relationships.

Acknowledgements

I would like to thank Dr. Katherine Maloney, whose support, knowledge and mentorship made my contribution to this project possible. Additionally, I thank my honors committee members, Dr. Matthieu Rouffet and Dr. Greg Crow for their suggestions and feedback. I would also like to thank the late Dr. Ryan Botts, without whom I never would have realized my interest in data analysis and whose loss was immensely felt while working on this project.

I am grateful to Dr. Allegra Aron and Dr. Mingxuan Wang who provided assistance with our LCMS data. Additionally, Dr. Gabriel Castro, who assisted us in using the LCMS instrument in the Jensen lab at UCSD. I would also like to thank the Coral Reef Research Foundation who provided resources for sample acquisition. Finally, thank you to both the Summer Undergraduate Research Fellowship program through the ACS organic division and Pfizer, Inc for providing funding.

References

(1) Farag, M.; Fekry, M.; Al-Hammady, M.; Khalil, M.; El-Seedi, H.; Meyer, A.; Porzel, A.; Westphal, H.; Wessjohann, L. Cytotoxic Effects of Sarcophyton Sp.. Soft Corals—Is There a Correlation to Their NMR Fingerprints? *Marine Drugs* **2017**, *15* (7), 211.

(2) France, S. C. What is a Sclerite? https://oceanexplorer.noaa.gov/okeanos/explorations/ex1905/logs/sept5/sept5.html (accessed Apr 7, 2022).

(3) Maloney, K. N.; Botts, R. T.; Davis, T. S.; Okada, B. K.; Maloney, E. M.; Leber, C. A.; Alvarado, O.; Brayton, C.; Caraballo-Rodríguez, A. M.; Chari, J. V.; Chicoine, B.; Crompton, J. C.; Davis, S. R.; Gromek, S. M.; Kurnianda, V.; Quach, K.; Samples, R. M.; Shieh, V.; Sultana, C. M.; Tanaka, J.; Dorrestein, P. C.; Balunas, M. J.; McFadden, C. S. Cryptic Species Account for the Seemingly Idiosyncratic Secondary Metabolism of Sarcophyton Glaucum Specimens Collected in Palau. Journal of Natural Products **2020**, 83 (3), 693–705.

(4) Nothias, L. F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; et al. *Bio Rxiv* **2019**, 812404. Feature Based Molecular Networking - GNPS documentation (https://ccms-ucsd.github.io/ GNPSDocumentation/featurebasedmolecularnetworking/)

(5) Sammarco, P.W., S. La Barre, J.C. Coll. Defensive strategies of soft corals of the Great Barrier Reef. *Oecologia* **1987**, *74,*93-101.

(6) Santacruz, L.; Thomas, O.; Duque, C.; Puyana, M.; Tello, E. Comparative Analyses of Metabolomic Fingerprints and Cytotoxic Activities of Soft Corals from the Colombian Caribbean. *Marine Drugs* **2019**, *17* (1), 37.

(7) Van Alstyne, K. L.; Wylie, C. R.; Paul, V. J.; Meyer, K. Antipredator Defenses in Tropical Pacific Soft Corals (Coelenterata: Alcyonacea). I. Sclerites as Defenses against Generalist Carnivorous Fishes. *The Biological Bulletin* **1992**, *182* (2), 231–240.

(8) Nothias, L. F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo, M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kameník, Z.; Kang, K. B.; Kessler, N.; Koester, I.; Korf, A.; Gouellec, A. L.; Ludwig, M.; Christian, M. H.; McCall, L.-I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.; Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nature Methods* **2020**.

Appendix

Mass Filtering		**ADAP Chromatogram Builder**		**Chromatogram Deconvolution**			
MS1 noise level	1.50E+03	Min group size in # of scans	3	Algorithm		Wavelets (ADAP)	
MS2 noise level	1.00E+01	Group intensity threshold	5.00E+03	m/z range for MS2 scan pairing (Da)		0.03	
		Min highest intensity	5.00E+03	RT range for MS2 scan pairing (min)		0.1	
		m/z tolerance	.02 m/z or 0 ppm				
Wavelets (ADAP) parameters		**Isotope peaks grouper**		**Join aligner**		**Peak filtering**	
S/N threshold	10	m/z tolerance	.05 m/z or 0 ppm	m/z tolerance	.05 m/z or 0 ppm	Area	none
S/N estimator	Intensity window	Retention time tolerance	0.15	Weight for m/z	25	Height	none
Min feature height	1000	Maximum charge	2	Retention time tolerance	0.05		
Coefficient/area threshold	25			Weight for RT	75		
Peak duration range	.01-2						
RT wavelet range	.01-.06						
Feature list rows filter			**Number of features**	4453			
Minimum peaks in a row	2						
Keep only peaks with MS2 scan							

Table A1: Detailed MZMine parameters for LCMS analysis.

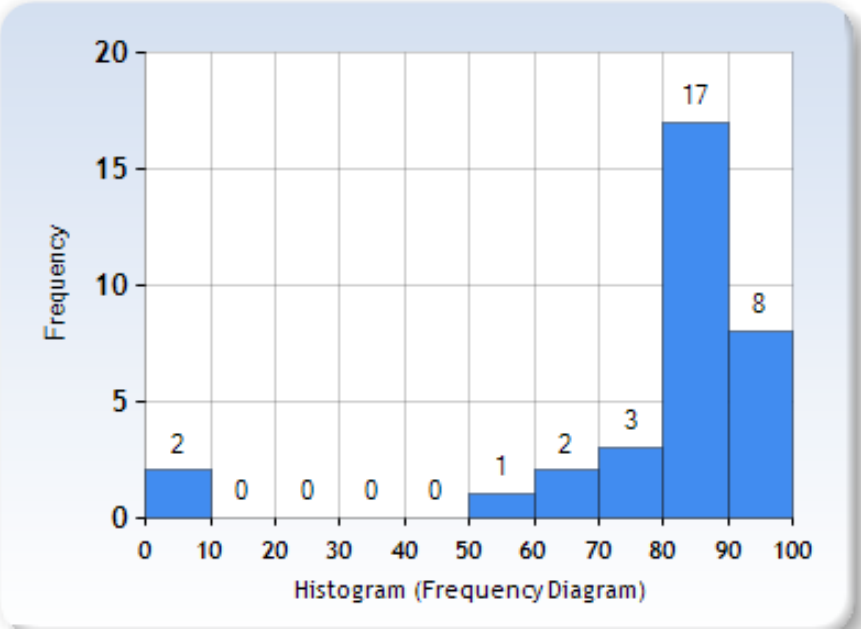


Figure A1: Histogram to determine sclerite abundance categories. Low percent mass = L = 0-60%, Medium percent mass = M = 61-80%, High percent mass = H = 81-90%, Very high percent mass = VH = 91-100%

Fraction A		Fraction B		Fraction C		Fraction G	
1H	13C	1H	13C	1H	13C	1H	13C
5.3	140.71	5.28	130.04	5.37	128.7	5.16	127
4.96	126.04	5.2	130.04	5.34	128.04	5.43	124.83
5.3	124.03	5.34	130.04	5.27	128.04	6.22	124.67
5.19	124.03	5.31	128.04	5.17	126.04	4.32	83.43
5.57	122.03	5.42	128.04	5.01	124.7	5.12	77.77
5.04	122.03	5.2	128.04	5.17	124.03	5.67	71.45
4.52	122.03	5.04	124.03	5.12	124.03	2.95	61.14
4.69	113.36	4.63	110.69	4.67	110.69	3.35	60.81
4.64	113.36	3.35	71.33	4.71	110.69	3.42	50.67
4.69	112.03	5.16	70	4.61	110.69	3.95	50.67
4.63	112.03	5.2	68.66	3.39	71.33	2.9	50.67
4.63	110.02	3.49	68.66	3.61	62.66	1.66	34.87
4.16	110.02	4	63.99	4.28	61.99	3.23	34.75
4.1	110.02	4.1	61.99	4.13	61.99	1.88	34.2
4.69	110.02	4.25	61.99	2.73	61.32	0.98	34.2
4.69	108.02	3.36	49.98	2.83	60.66	2.06	34.04
4.63	108.02	3.41	49.98	3.46	49.98	2.96	33.87
4.69	106.69	2.12	39.31	2.62	40.64	2.52	33.04
4.64	106.69	2.31	33.97	1.99	36.64	1.2	29.05
3.48	50.65	2.46	33.97	1.46	36.64	1.83	27.05
2.57	41.31	2.25	33.97	2.05	35.31	1.53	27.05
2.54	39.31	1.08	29.3	2.24	35.31	1.34	23.89
2.53	38.64	1.27	28.63	2.31	33.3	2.08	20.9
2.08	38.64	2.02	26.63	1.94	31.3	1.33	18.24
1.76	33.3	2.74	25.3	1.31	31.3	1.57	17.08
1.91	33.3	2.84	25.3	1.2	29.3	1.11	16.91
1.91	31.97	1.57	24.63	1.38	28.63	1.2	16.41
1.19	29.3	2.12	23.96	1.31	28.63		
1.25	29.3	0.73	13.29	1.54	25.97		
1.69	27.97	0.83	13.29	1.45	25.97		
1.34	27.97			2.75	25.3		
0.98	27.97			2.83	25.3		
2.22	24.63			1.59	23.96		
2.27	24.63			1.39	23.96		
2.1	23.3			2.16	21.96		
0.81	19.29			2.08	21.96		
1.69	18.63			0.81	19.96		
1.61	18.63			1.66	19.96		
1.54	17.29			1.63	15.29		
1.56	15.29			1.59	15.29		
0.87	13.96			0.92	13.29		
				0.85	13.29		
				0.64	11.29		

Table A2: HSQC data for fractions A, B, C and G used to generate SmartNMR tables (condensed into 1 spreadsheet).

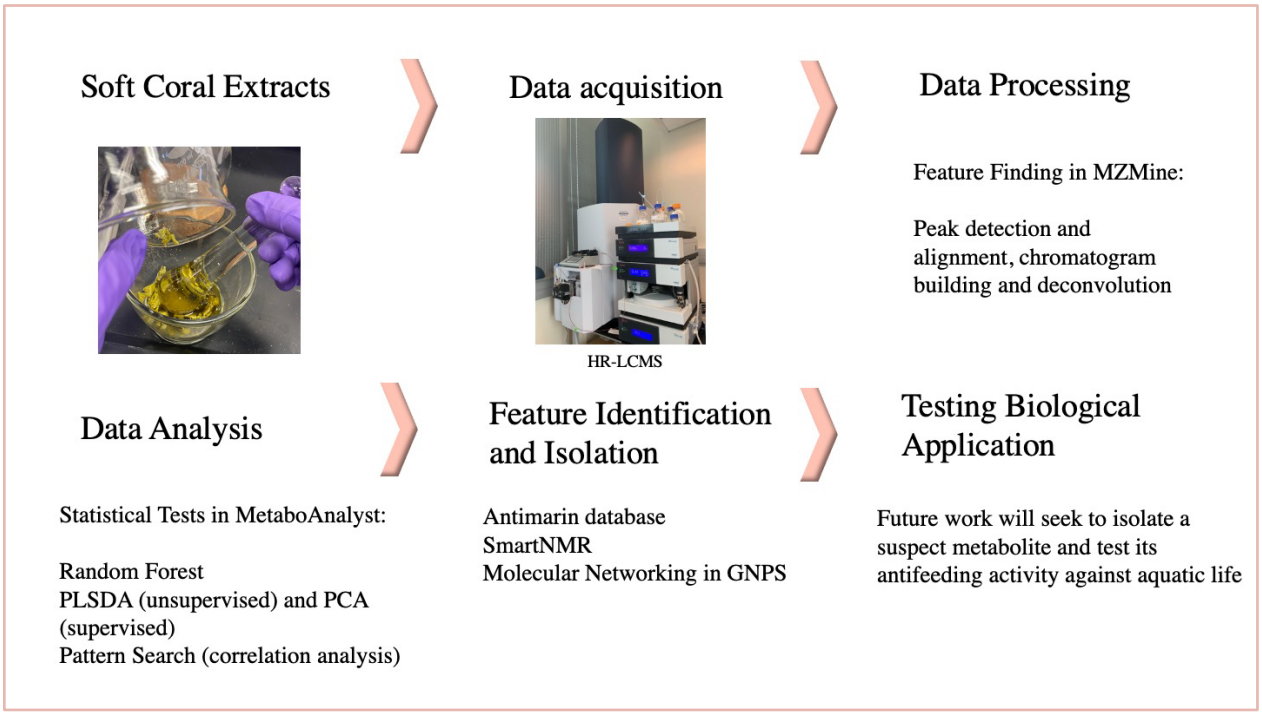


Figure A2: Project workflow summary.