Title
Functional Annotation of Natural Products: Methods and Applications to Marine Microbiota

Permalink
https://escholarship.org/uc/item/8498h1mk

Author
Ochoa, Jessica Leigh

Publication Date
2017

Peer reviewed|Thesis/dissertation
FUNCTIONAL ANNOTATION OF NATURAL PRODUCTS:
METHODS AND APPLICATIONS TO MARINE MICROBIOTA

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Jessica L. Ochoa

June 2017

The Dissertation of Jessica L. Ochoa
is approved:

_____________________________
Professor Roger Linnington, Advisor

_____________________________
Professor Joseph Konopelski, Chair

_____________________________
Professor Rebecca Braslau

_____________________________
Tyrus Miller
Vice Provost and Dean of Graduate Studies
# Table of Contents

1. Phenotype-Guided Natural Products Discovery Using Cytological Profiling ........................................ 1

1.1. Introduction ......................................................................................................................... 4

1.2. Results and Discussion .................................................................................................... 8

   1.2.1. Analysis of Crude Bioactives ...................................................................................... 8

   1.2.2. Peak Library Generation .............................................................................................. 12

   1.2.3. Compound identification .............................................................................................. 14

   1.2.4. Pure Compound Screening ............................................................................................ 15

   1.2.5. Cluster Analysis ............................................................................................................ 19

1.3. Conclusion ....................................................................................................................... 22

1.4. Experimental ..................................................................................................................... 23

   1.4.1. General Experimental Procedures ............................................................................... 24

   1.4.2. Fermentation and Isolation ......................................................................................... 24

   1.4.3. Cytological Profiling and Screening ............................................................................ 25

   1.4.4. Synthesis of XR344 ..................................................................................................... 26

   1.4.5. Co-injection of synthetic and natural product XR334 (3) ........................................ 28

   1.4.6. $^1$H NMR (600MHz, CDCl$_3$) spectrum of synthetic and isolated XR334 .......... 29

   1.4.7. $^{13}$C NMR (150 MHz, CD$_3$OD) spectrum of XR334 (3) ....................................... 30

   1.4.8. $^1$H NMR (600MHz, CD$_3$OD) spectrum of nocapyrone B (4) ............................ 31
1.4.9 $^1$H NMR (600MHz, CD$_3$OD) spectrum of nocapyrone H (5).......................... 32

1.4.10 $^1$H NMR (600MHz, CD$_3$OD) spectrum of nocapyrone L (6).................... 33

1.4.11. Expanded CP fingerprint of nocapyrone L (6) ........................................... 34

1.4.12. HPLC trace for reverse phase stage fractionation of RLUS1665D ............. 35

1.4.13. Dilution Series of XR334 and Nocapyrone L ............................................. 36

1.5. References ........................................................................................................... 37

2. The Marine Mammal Microbiome Yields a Novel Antibiotic with Potent Activity Against
Clostridium difficile ......................................................................................................... 45

2.1. Introduction ............................................................................................................ 46

2.2. Results and Discussion .......................................................................................... 49

  2.2.1. Sample Collection and Isolate Prioritization for Novel Bioactives Discovery.. 49

  2.2.2. Biological Screening of Crude Extracts. ......................................................... 51

  2.2.3. Structure Elucidation. ....................................................................................... 52

  2.2.4. Biological Activity ............................................................................................ 56

2.3. Conclusion ............................................................................................................. 61

2.4. Experimental ......................................................................................................... 63

  2.4.1. General Experimental Procedures. ................................................................. 63

  2.4.2. Cultivation of Bacteria..................................................................................... 63

  2.4.3. DNA Isolation, PCR Amplification and Sequencing. ..................................... 64

  2.4.4. Phylogenetic Analysis. .................................................................................... 65
2.4.5. Molecular Networking Experimental: ................................................................. 69
2.4.6. Extraction of Cultivated Isolates. ........................................................................ 70
2.4.7. Fermentation and Isolation. .................................................................................. 71
2.4.8. Flow Cytometry. .................................................................................................. 72
2.4.9. Bacterial Strains. .................................................................................................. 73
2.4.10. High-Throughput Antibacterial Inhibition Assay .............................................. 73
2.4.11. Cytological Profile Screening and Image Analysis. ........................................... 75
2.4.12. Extended Structure Elucidation ......................................................................... 77
2.4.13. Configurational Analysis .................................................................................... 82
2.4.14. NMR Chemical Shifts ....................................................................................... 87
2.4.15. NMR Spectra ..................................................................................................... 88

2.5. References .................................................................................................................. 99

3. Mining Genetic Interactions in *Saccharomyces cerevisiae* for Drug Discovery .......... 107

3.1. Introduction ................................................................................................................ 109

3.2. Results/Discussion ..................................................................................................... 116

3.2.1. Prioritization by betweenness centrality ............................................................... 116
3.2.2. Validation using Boone data .............................................................................. 119
3.2.3. Binning of genes using Gene Ontology .............................................................. 122
3.2.4. “Hub gene” selection and ranking ...................................................................... 126
3.2.5. Comparison of acquired data to published data ................................................. 131
3.3. Conclusion ................................................................................................................. 133

3.4. Experimental .............................................................................................................134

  3.4.1. Cell culturing ........................................................................................................134

  3.4.2. Data Analysis .......................................................................................................134

3.5 References ............................................................................................................... 137
# Table of Figures

## Chapter 1

Figure 1.1 Cytological profiling of 1665D ................................................................. 11

Figure 1.2 Clustering of pure compounds ................................................................. 16

Figure 1.3 Structures of known antimitotic agents and compounds isolated from RLUS1665 ......... 17

Figure 1.4 Gene sets enrichment analysis (GSEA) ...................................................... 21

## Chapter 2

Figure 2.1 Overview of project workflow ................................................................. 50

Figure 2.2 Crude extract BioMAP dilution profiles ...................................................... 52

Figure 2.3 Structure elucidation of phocoenamicin .................................................... 53

Figure 2.4 Structure of phocoenamicin, closest analogue maklamicin, and fidaxomicin .............. 55

Figure 2.5 Cytological profiling of phocoenamicin .................................................... 59

Figure 2.6 Membrane polarity and permeability assays ............................................. 61

## Chapter 3

Figure 3.1 Synthetic Lethality ................................................................. 114

Figure 3.2 Representation of betweenness centrality ................................................. 117

Figure 3.3 Yeast interaction network ................................................................. 117

Figure 3.4 Screening results ................................................................. 118

Figure 3.5 Clustering of 50 genes with highest betweenness centrality scores .............. 120

Figure 3.6 Selected sections from the clustering of control compounds .............. 121
Figure 3.7 Gene ontology network for *S. cerevisiae*.............................................................122

Figure 3.8 Distributions of GO terms throughout the genome.............................................123

Figure 3.9 Idealized screening results..................................................................................125

Figure 3.10 NxN matrix of the top ten GO terms.................................................................126

Figure 3.11 NxN matrix of genes binned by GO terms..........................................................127

Figure 3.12 Heatmap of top DNA damaging genes...............................................................129

Figure 3.13 Enriched GO terms............................................................................................130

Figure 3.14 Comparison of experimental data with published data....................................132
Abstract

Functional Annotation of Natural Products:

Methods and Applications to Marine Microbiota

By

Jessica L. Ochoa

Natural products have made dramatic contributions to a significant portion of FDA approved therapeutics. While many of these secondary metabolites have been exhaustively studied, the true purpose of these molecules is often left unresolved. Mechanism of action (MOA) annotation remains one of the largest bottlenecks in drug discovery. In addition, the source of natural products affects chemical diversity and metabolites isolated from microbiota are an emerging source of antimicrobials. They are believed by many to modulate host health, affecting predisposition to disease and pathogen invasion. Through the application of unbiased screening methods to novel sources, a more comprehensive understanding of the global effects of natural products can be achieved. This dissertation will present the implementation of an image based cytological profiling platform for the isolation of compounds with a specific, predefined mode of action, while simultaneously characterizing all bioactive constituents isolated from the crude extract. The development of a training set independent genetic assay in *Saccharomyces cerevisiae* will be discussed. In addition, the application of image based and genetic techniques to a new compound,
phocoenaminicin, isolated from marine mammal microbiota demonstrates the importance of identifying both new sources of natural products while simultaneously developing new tools to characterize these compounds, generating a more comprehensive image of what natural products are and what they are capable of.
Acknowledgements

This dissertation would not have been possible without the support of my family, friends and colleagues. I would like to thank my husband Tony Brock and Heidi Ochoa for their ceaseless love and support during my graduate career, they have given me so much joy and an essential work-life balance. My parents, Don and Ardis, have provided me unconditional encouragement and love my entire life and have always inspired me to work hard and value education. Todd and Heather, my siblings and my best friends growing up, have always been such a special treasure to have and they have blessed me with unlimited silliness and love as well as our new family members Leah, Cameron and Aaron who have become very precious to us all. My grandmother Barbara has shown me what true independence and strength is and has become one of my dearest friends over the years. My family in-law Debbie, Ernie, Niki, Mike, June, Alice and Hazel have never ceased to surprise me with how enthusiastically they have welcomed me into their family. Walter Bray has also been a constant friend and mentor, taking the time to help me at every step of my career and has taught me not only how to rigorously design an experiment with proper controls, but also what true generosity looks like. The Linington lab family has also been an unexpected yet wonderful family that has ignited lifelong friends. Laura is my big sister in science, she gave me an amazing research project, and has taken special care to mentor me throughout my career, shaping me as a scientist even though we only overlapped a few months in lab. Kenji has also been an incredible mentor and the graduate student that I have worked alongside the longest. His work ethic, brilliance, and dedication to mentorship is...
unrivaled and Kenji all I have left to tell you is: we made it!! Kelly, Chris S., Gabe, Chris W., Bailey, Jake, Weng, and Cat have all taken the time to critique my scientific progress and have all taught me very different lessons in and outside of the Linington lab that I am very grateful for. I learned how to solve complex structures with NMR from Roger and that has become my greatest joy as a scientist, and I thank him for recommending me for a talk in the Gordon Research Conference which has been the brightest moment of my scientific career.

Losing the scientific support, resources and emotional support of Linington lab when they relocated to Canada during the most critical point of my graduate career ended up being much more challenging than I could have ever anticipated. I am still amazed and forever grateful for the generosity that Scott Lokey and his lab members have shown me. They have welcomed me whole heartily into their lab and shared their recourses, time and mentorship with me as freely as if I had already been a part of their lab for years. Scott helped design two of my thesis projects and Josh, Andy, Dustin, Cameron and Walter helped me execute them scientifically after the lab departed. Alicia is one of the few female graduate students in chemistry and I am continually amazed and humbled by her success and perseverance. My committee members Rebecca Braslau and Joe Konopelski have been an unwaivering source of mentorship and support and have continually pushed me to be a better scientist. I came to UCSC for graduate school because on my first visitation it was clear that people in this department truly cared for each other and were genuinely invested in each other’s success, I have been lucky enough to experience that compassion first hand.
The text of this dissertation includes reprints from the following published material:


The text of this dissertation includes data from the following manuscripts in preparation:


The co-authors listed in these publications and manuscripts directed, supervised, or contributed to the research which formed the basis for this dissertation.
Chapter 1

Phenotype-Guided Natural Products Discovery Using Cytological Profiling
Table of Contents

1. Phenotype-Guided Natural Products Discovery Using Cytological Profiling ..... 4
   1.1. Introduction ........................................................................................................ 4
   1.2. Results and Discussion ....................................................................................... 8
      1.2.1. Analysis of Crude Bioactives ....................................................................... 8
      1.2.2. Peak Library Generation ............................................................................. 12
      1.2.3. Compound identification ........................................................................... 14
      1.2.4. Pure Compound Screening ......................................................................... 15
      1.2.5. Cluster Analysis .......................................................................................... 19
   1.3. Conclusion .......................................................................................................... 22
   1.4. Experimental ...................................................................................................... 23
      1.4.1. General Experimental Procedures ............................................................. 23
      1.4.2. Fermentation and Isolation .......................................................................... 24
      1.4.3. Cytological Profiling and Screening ............................................................ 25
      1.4.4. Synthesis of XR344 ................................................................................... 26
      1.4.5. Co-injection of synthetic and natural product XR334 (3) ......................... 28
      1.4.6. \textsuperscript{1}H NMR (600MHz, CDCl\textsubscript{3}) spectrum of synthetic and isolated XR334 .......... 29
      1.4.7. \textsuperscript{13}C NMR (150 MHz, CD\textsubscript{3}OD) spectrum of XR334 (3) ........................ 30
      1.4.8. \textsuperscript{1}H NMR (600MHz, CD\textsubscript{3}OD) spectrum of nocapyrone B (4) ....................... 31
      1.4.9. \textsuperscript{1}H NMR (600MHz, CD\textsubscript{3}OD) spectrum of nocapyrone H (5) ....................... 32
      1.4.10. \textsuperscript{1}H NMR (600MHz, CD\textsubscript{3}OD) spectrum of nocapyrone L (6) ......................... 33
      1.4.11. Expanded CP fingerprint of nocapyrone L (6) ........................................... 34
      1.4.12. HPLC trace for reverse phase stage fractionation of RLUS1665D ............. 35
1.4.13. Dilution Series of XR334 and Nocapyrone L................................................... 36

1.5. References.................................................................................................................. 37
1. Phenotype-Guided Natural Products Discovery Using Cytological Profiling

1.1. Introduction

Although natural products have proven value in cancer chemotherapeutic development, the process from the original collection to the biological annotation of natural products is labor intensive and time consuming. For example, the microtubule stabilizing activity of the natural product dictyostatin 1 was not established until 10 years after it was originally isolated.\(^1\) In addition, collection of the plant material containing paclitaxel occurred in the 1960s but its microtubule stabilizing activity was not published until 1979.\(^2\) Traditionally, targeted anticancer discovery for natural products is performed by screening crude extracts in high-throughput screens using reporter assays such as enzyme-linked immunosorbent assays (ELISAs) or reporter gene assays (RGAs). While these assays are high-throughput, they typically only report on a single molecular target or pathway, meaning that discovery is limited to a specific set of biological targets.\(^3\) These population measurements miss variations in individual cell dynamics and morphology, and hits in these screens are therefore unable to account for off-target affects, and are prone to false positives when broadly cytotoxic compounds are encountered. For example, screening campaigns for a selected target such as a specific enzyme can yield hits that are indistinguishable from compounds that are also active against a wider set of targets and whose development potential is therefore typically low.

Tremendous progress has been made in recent years to develop high-content imaging systems capable of quantifying multiparametric cellular responses of drug
treated cells. These data can be viewed as phenotypic “fingerprints” which have been successfully used to differentiate classes of drugs with discrete modes of action (MOAs). This screening approach takes a global view of biological attributes of bioactive compounds by determining the overall effect of small molecules on cell morphology, rather than examining specific molecular targets or pathways as is common in target-based screening. While high-content screening (HCS) is an extremely powerful tool, it is predominately used for pure compound analysis; therefore the time between sample collections and MOA annotation still exists as a bottleneck in drug discovery.

Development of untargeted HCS of natural product extracts is providing a mechanism to circumvent this bottleneck by obtaining detailed mechanistic information for large natural product libraries in a high-throughput fashion. Elimination of lengthy purification and workup protocols for non-bioactive or broadly cytotoxic compounds rapidly prioritizes compounds with desired bioactivities. Since the emergence of automated image analysis, the development of algorithms that process and categorize images to predict MOAs has provided new opportunities for the analysis of compounds without published MOAs. In our effort to expedite hit-to-lead development from natural product libraries, we have recently reported the implementation of a high-content cytological profiling platform that uses image-based screening to directly visualize the phenotypic effects of natural product extracts on HeLa cell development. These previous reports analyzed the major compound clusters found from screening a pilot library of 312 extracts to validate this screening
tool and identified both known and novel natural products. In this current study. This technology was applied to our entire 5304-member extract library, and these data were used to test the hypothesis that cytological profiling (CP) could be used to discover compounds with specific and predefined modes of action.

Several approaches have been employed to derive multiparametric annotation for natural product extracts. Predominately applied to pure compound libraries, untargeted profiling utilizes bioactivity profile matching where compounds with similar phenotypic responses are inferred to have similar mechanisms of action. Among these, gene expression profiling and yeast gene deletion libraries have been used successfully to annotate natural product MOAs in an untargeted manner.\textsuperscript{10,11} Zebrafish imaging has also been utilized to characterize whole animal response to natural products and is one of the few whole organism approaches to untargeted natural products screening.\textsuperscript{12} A number of computational tools exist to aid in the interpretation of image-based screening data; however, to date there have been few examples of the application of these methods to unbiased natural products discovery.

To test the hypothesis that high-content screening can be used to discover bioactive compounds with pre-specified modes of action, our library was explored for extracts displaying an antimitotic phenotype, and this demonstrated that this bioactivity profiling approach can be used to map a single specific constituent within an extract to this predefined biological activity. This proof of concept provided evidence that untargeted image-based screening can be used effectively for the discovery of compounds with specific mechanisms of action and opens the door for
the use of this technology to discover compounds that target other high-priority pathways and molecular targets from crude natural products in a high-throughput manner.

Induction of mitotic arrest is a validated anticancer mechanism, with tubulin being one of the most studied cancer targets, and the target of a number of chemotherapeutic agents.\textsuperscript{13} However, many mitosis-targeting treatments currently struggle with incomplete neoplasm eradication, resistance and toxicity issues.\textsuperscript{14} Recent antimitotic drug discovery has focused on analyzing specific aspects of mitosis, often accomplished by screening compounds in tubulin binding assays or in whole cell assays where the amount of mitotic arrest with respect to vehicle treated cells is analyzed. As an example, a high throughput ELISA assay has been used to discover new paclitaxel analogues with similar activity to the parent compound.\textsuperscript{15} In addition, biphenabulin, an antimitotic agent with nanomolar activity, was recently discovered utilizing whole cell imaging of cells stained with phosphorylated histone H3 (pHH3), a mitotic marker, to screen libraries of synthetic compounds created by diversity orientated synthesis of macrocycles with natural-product-like geometry.\textsuperscript{16} Our new study focuses on determining whether our unbiased CP screen could be utilized to identify compounds that affect tubulin dynamics from complex natural product extracts, while still providing detailed yet untargeted biological annotations for all bioactive constituents.
1.2. Results and Discussion

1.2.1. Analysis of Crude Bioactives

In the original report describing the development of cytological profiling for natural products discovery a pilot library of 312 extracts was evaluated and the bioactive constituents were characterized from a subset of the bioactive fractions. This analysis has now been extended to our full natural products library, consisting of 5304 prefractionated extracts from marine-derived Actinobacteria. Cytological profiling of this entire library, followed by hierarchical clustering with the training set of 480 compounds from the Harvard Institute of Chemistry and Cell Biology collection (ICCB) possessing known MOAs revealed 41 discrete clusters, including 27 clusters which contained both bioactive extracts and reference compounds from the training set. Among these there are several expanded versions of clusters that were originally identified and annotated in our initial cytological profiling paper, including DNA synthesis inhibitors, vacuolar ATPase inhibitors, potassium channel inhibitors, and kinase inhibitors. In addition, this expanded cytological profiling set contains multiple clusters not observed in the original training set, many of which contain training set compounds with related MOAs.

Finally, there are numerous smaller clusters that do not cluster well with compounds from the training set, suggesting either that these extracts contain compounds with MOAs not represented by members of the training set or that these extracts contain multiple bioactive constituents that cause “mixed-mode” phenotypes that are not representative of any of the individual components (an inherent limitation
of cytological profiling with complex mixtures). One of the crude extracts revealed the presence of borrelidin, a potent inhibitor of bacterial and eukaryotic threonyl-tRNA synthetases. Upon further fractionation, multiple bioactives were identified and a new analogue, borrelidin B was discovered without the presence of the nitrile moiety present in the original borrelidin analogue. Screening of the two compounds in an assay for tRNA synthetase inhibition suggested that the nitrile was essential for activity. While inactive in the tRNA assay, borrelidin B was active in our cytological profiling assay displaying the versatility of this untargeted approach.

Among the remaining active clusters, one small subset clustered tightly with the known microtubule poisons nocodazole and paclitaxel, suggesting that these extracts contain compounds that disrupted tubulin dynamics. In general, CP fingerprints of known microtubule stabilizers and destabilizers cluster closely together but can be differentiated by comparing tubulin staining from DMSO controls. Tubulin stabilizers such as paclitaxel show a positive deviation while destabilizers such as colchicine and nocodazole exhibit a negative deviation in tubulin staining (Figure 1.1A). Analysis of the heatmap suggested that the clustering was predominately driven by an increase in the mitotic index. Inspection of the well images for extract RLUS1665D (Figure 1.1A) showed a strong mitotic stall and diminished tubulin staining consistent with the phenotypes observed for neighboring reference compounds that inhibit tubulin polymerization (Figure 1.1B). The precise order of extracts and reference compounds in the hierarchical clustering network is contingent upon both the number and concentration of compounds in the original
input set. Clustering these active extracts with several different dilution series for the training set of known bioactives identified extract RLUS1665D as having the most reliable clustering with known antimitotic agents. We therefore prioritized RLUS1665D for further chemical evaluation.
Figure 1.2 Cytological profiling of 1665D. (A) Fluorescence images of drug treated wells all show deviations in tubulin (green) and mitotic cells (cyan) from the control. The corresponding phenotypic fingerprint generated from the image analysis is shown above each well image. (B) Cytological profiling fingerprint of original MNP cluster and microtubule poisons. Positive deviations from DMSO treated wells are displayed in yellow, negative deviations are displayed in blue. Features highlighted for tubulin and mitotic cells display an increase in mitotic cells and variable tubulin staining.
1.2.2. Peak Library Generation

Our standard library preparation protocol involves organic extraction of microbial culture broths, followed by initial prefractionation on C$_{18}$ SPE cartridges using a MeOH/H$_2$O step gradient to generate seven prefractions for screening.$^{4,5,8}$ An aliquot of prefraction RLUS1665D was first examined using our standard peak library separation protocol.$^{6,8}$ In brief, extracts are analyzed by RP-HPLC using an analytical fraction collector to separate the eluent into deep well 96-well plates in one minute intervals (Phenomenex Synergi Fusion-RP 10 micron, 80 Å, 250 x 4.6 mm, 25:75% MeCN/H$_2$O + 0.02% formic acid to 90:10% MeCN/H$_2$O + 0.02% formic acid; 2 mL min$^{-1}$ flow rate). A portion of this eluent (~5%) is diverted from the fraction collector to an ESI-single quadrupole mass spectrometer performing fast polarity switching to simultaneously acquire mass spectra in both positive and negative ionization modes. Together these analyses provide retention time, UV absorbance and mass spectrometric data for all constituents from each extract. When combined with screening data from the eluent plate these data can be used to connect individual constituents with observed bioactivities from primary screening results.

In the case of RLUS1665D, the peak library revealed the presence of two major families of compounds, both of which showed weak phenotypes in the cytological profiling screen of the peak library. One limitation of peak library screening is that if an extract contains families of compounds with similar bioactivities but different retention times, separation into individual wells can decrease the titer in each well sufficiently to reduce the strength of the observed
phenotype. In this instance, refermentation of the producing organism followed by a combination of normal phase silica gel and reversed phase HPLC steps provided fractions containing these two compound families, each of which displayed strong independent cytological profiles when reevaluated in the cytological profiling screen.

Examination of the screening results from the second round of fractionation indicated that one of these compound families recapitulated the original antimitotic phenotype, while the other family was strongly cytotoxic at the tested concentration, affording test wells with few cells (experimental 1.4.12). This result highlights the second challenge with phenotypic screening, whereby testing at too high a concentration leads to a situation where few cells remain for image analysis. This “death phenotype” affords cytological profiles where the majority of size and shape features are scored as strongly negative (due to the low cell count), which in turn precludes accurate MOA predictions. This limitation is resolved by screening test materials as fine-scale dilution series (typically 16 two-fold dilutions with a concentration range of 100 µM – 3 nM). Appropriate concentrations for clustering are then selected by identifying the first dilution with a cell count within three standard deviations of the mean cell count for untreated control wells. In this instance, final purification by sequential stages of RP-HPLC chromatography led to the isolation of representatives from both compound classes. Screening of these pure compounds as dilutions series in the CP assay (experimental 1.4.13) confirmed that one of these compounds displayed an antimitotic cytological profile (Figure 1.2A), while the other clustered tightly with the known calcium channel modulators.
benzamil, BAY-K-8644, and A-23187 (Figure 1.2B and C). This result highlights how cytological profiling can be used to both focus on the discovery of compounds within a specific mechanistic class (antimitotic agents) while simultaneously annotating compounds with unknown MOAs (calcium channel modulators).

### 1.2.3. Compound identification

The identity of the antimitotic agent was determined through a combination of 1D- and 2D-NMR analyses, mass spectrometry and total synthesis. Initial ESI-single-quadrupole LCMS analysis identified two mass spectrometric features consistent with the [M+H]$^+$ and [M-H]$^-$ adducts with $m/z$ values of 321.2 and 319.0 respectively. Subsequent HRESITOFMS analysis revealed an [M+Na]$^+$ adduct with $m/z$ 343.1875, consistent with the molecular formula C$_{19}$H$_{16}$N$_2$O$_3$. Examination of the UV-absorbance spectrum for this compound revealed strong absorbances at 250 and 352 nm, suggestive of a conjugated aromatic compound. A query of the Antimarin database revealed one compound, diketopiperazine XR334 (Figure 1.3), which possessed a matching molecular formula, as well as identical NMR and UV spectroscopic features. Because this scaffold is synthetically accessible through a two-step condensation reaction between diacetyl-2,5-piperazinedione, benzaldehyde and $p$-anisaldehyde, the structure of the isolated antimitotic agent was confirmed by total synthesis (see experimental 1.4.4), and validated by NMR and HPLC-MS comparison with the original isolated material (see experimental 1.4.5-6).
1.2.4. Pure Compound Screening

In addition to the structural verification, the biological activity for the isolated natural product was confirmed through parallel CP screening for dilution series of both the natural and synthetic materials (Figure 1.2A). These results demonstrated that the synthetic and naturally occurring material possessed the same cytological profiles, and confirmed that the observed activity was caused by XR334, rather than any potent but minor constituents present as contaminants in the natural sample. Clustering of the CP fingerprints from the synthetic and natural samples of XR334 with the training set library revealed that XR334 clustered closely with other microtubule poisons, including nocodazole and vinblastine (Figure 1.2B).
Figure 1.3 Clustering of pure compounds. (A) Cytological profiling analysis of isolated XR334 (XR334_NP) and synthesized XR334 (XR334_SYN). (B) Cytological profiling analysis of nocapryrone L. (C) Chemical structures of compounds that clustered with nocapryrone L.
XR334 is structurally similar to plinabulin (2), a synthetic analogue of the natural product phenylahistin (1) (Figure 1.3). Plinabulin is both a vasculature-disrupting agent and antimicrotubule agent that maintains efficacy \textit{in vivo} and works synergistically with docetaxel, a synthetic derivative of paclitaxel.\textsuperscript{13,19} Examination of the high-content images generated in the CP screen showed that XR334 displays moderate tubulin depolymerization compared to DMSO controls or strong destabilizers such as nocodazole. Therefore, while this clustering is predominately driven by the increased mitotic index, closer inspection of its effect on the microtubule cytoskeleton morphology indicates that it is possible to further subdivide this large cluster of microtubule poisons into several classes, based on their effect on tubulin polymerization. This is consistent with MOA studies performed on synthetic analogues of plinabulin.\textsuperscript{14,24-26}

While screening approaches focused on

![Figure 1.4 Structures of known antimitotic agents and compounds isolated from RLUS1665](image)
antimitotic compounds would have successfully identified XR334 as the bioactive constituent, upon peak library generation they would be unable to annotate compounds within the extract with alternative MOAs. In addition to XR334, a second class of bioactive compounds was also identified from this extract. This second class of compounds shared similar UV profiles, and differed by sequential mass differences of 14 Da, consistent with variation in the number of methylene units in a set of structurally related analogues. Initial isolation of one of these family members afforded a white solid that gave a HRMS [M+H]$^+$ signal of 267.1957 consistent with the molecular formula $\text{C}_{16}\text{H}_{26}\text{O}_{3}$. $^1$H and $^{13}$C NMR analyses revealed the presence of one carbonyl, one methoxy, and four methyl groups. Further NMR analysis identified nocapyrone L (Figure 1.3) as a candidate match, which was subsequently confirmed by comparison of NMR and MS literature data.$^{15,27}$ The remaining pyrone analogs from this series were identified as nocapyrones B (4) and H (5) through comparison of methylation patterns of $^1$H and COSY NMR spectra versus nocapyrone L and comparison with previously reported literature data.$^{16,27-29}$

$\gamma$-Pyrones have been identified in a number of screening campaigns,$^8,28,29$ and have previously been reported to impact intracellular calcium concentrations.$^{17,27}$ To predict the MOA of these compounds against HeLa cells, a dilution series of the most active compound, nocapyrone L (6), was evaluated in the CP assay. Nocapyrone L was isolated previously from a venom duct of a sea snail, Conus rolani, and was found to be active in a calcium imaging assay of dissociated dorsal root ganglion neurons from.$^{27}$ In line with previous reports, sublethal concentrations of compound 6
clustered closely with compounds that affect calcium channels from the ICCB library training set including Bay K-8644 and benzamil (Figure 1.2B). These results indicate that disruption of calcium channel function is likely the predominant MOA for the observed cytotoxicity of the nocapyrones in our assay system. This result demonstrates that the CP platform can differentiate between a wide array of biological mechanisms, even if these mechanisms are not directly reported on by the structural and cell cycle stains used in CP analysis.

1.2.5. Cluster Analysis

In order to determine which cytological phenotypes underlie the clustering of the nocapyrones with the calcium channel modulators from the reference library, we used a modification of gene sets enrichment analysis (GSEA) using software publically available from the Broad Institute.\textsuperscript{10} GSEA was originally developed to identify gene sets (i.e. sets of genes grouped by function, biological process, co-regulation, etc.) that are enriched in genes whose up or down regulation are correlated with a particular phenotype (e.g. metastatic vs. non-metastatic) among a set of samples (e.g. tissue samples from different patients).\textsuperscript{30,31} For example, initial gene analysis of type 2 diabetes mellitus (DM2) revealed that no single gene could be correlated to biopsy samples taken from patients with DM2. While the expression of individual genes can have large variations within sample sets, if these genes are placed in gene sets based on their connection to specific biological pathways it is then possible to identify consistent correlations between gene set expression and specific
phenotypes. Analysis of DM2 tissue samples using GSEA revealed the correlation between the expression of OXPHOS-CR genes and DM2.\textsuperscript{31}

To apply this approach to the chemical genetic signatures derived from the cytological profiling screen, we grouped the CP features into 23 classes (analogous to the gene sets discussed above) based on the broad phenotypic signatures on which they report (e.g., nuclear shape, pHH3 intensity in large areas, cell count, mitotic index, etc.). Applying GSEA to our dataset, we replaced genes with CP features, gene sets with feature classes, phenotypes with annotated MOA classes, and samples with compound-dose instances. This allowed one to ask, for the calcium ion channel modulators and nocapyrones that cluster with them, which CP feature classes correlate significantly with the distinction between these compounds and the rest of the library. The advantage of this analysis is that it provides a quantifiable method for identifying key biological features that drive associations between test compounds in the cytological profiling clusters, even if these features are not readily observable by eye. Knowing which features are drivers of clustering for a specific group of compounds provides an additional tool for evaluating the biological significance of a given cluster by highlighting specific size and shape cell deviations that are positively correlated with the clustering of any specific group of compounds.

This “GSEA-CP” analysis identified enrichment in nuclear stain 5-ethynyl-2’-deoxyuridine (EdU) intensity features among those that correlated with the cluster containing the nocapyrones and calcium ion channel modulators (Figure 1.4). The EdU stain is a modified nucleoside analogue that reports on cells actively undergoing
DNA synthesis. For each of the feature classes, an enrichment score was generated. The enrichment score reflects the significance of a feature class in terms of how well its individual members are correlated with a designated phenotype. Expansion of these features enriched in this class for nocapyrone L (6) identified that six out of the seven enriched features from the GSEA analysis show significant positive deviation from the DMSO control (see experimental 1.4.11). Visual inspection of the images did not reveal this subtle phenotypic difference, highlighting the utility of computational image processing in classifying compounds by phenotype. Therefore, CP is a powerful tool capable of deciphering predetermined mechanisms from complex data sets, yet flexible enough to annotate MOAs of compounds with orthogonal bioactivities within the same extract.

Figure 1.5 Gene sets enrichment analysis (GSEA). Analysis of EdU features for calcium channels and nocapyrone L. Seven EdU features were identified to drive the clustering of this class as denoted by asterisks; six of these enriched features showed significant positive deviations for the cells treated with nocapyrone L as denoted by black asterisks.
1.3. Conclusion

This study demonstrates the utility of untargeted morphology-based screening for the broad classification of natural product MOAs. By quantifying specific morphological features of HeLa cell development under sublethal drug pressure, the CP platform is capable of distinguishing a large number of discrete phenotypes that in turn report on the pathways and mechanisms disrupted by drug treatment. Targeted phenotype matching from untargeted cytological profiling data can be used to identify compounds with specific predefined MOAs. In this case, the antimitotic XR334 was directly targeted and identified from our complex natural product library through a combination of CP screening and analytical chemistry.

The parallel identification of nocapyrones B (4), H (5) and L (6) as predicted calcium ion channel modulators demonstrates that the CP platform can be used to discover and characterize bioactive constituents whether these bioactive compounds directly target processes reported on by the staining set (e.g. XR334, which disrupts tubulin dynamics) or not (e.g. nocapyrone L, which disrupts calcium channel function). These results therefore suggest that unbiased phenotypic profiling may be a valuable tool for the broad characterization of bioactive constituents from natural product libraries, and that this approach can be employed to identify compounds with specific biological MOAs by targeting relevant clusters in the CP profiles. In this study, the practicability of this approach has been demonstrated by mining our natural product libraries for compounds with specific mechanisms, and validating their
predicted MOAs through a combination of NP isolation, synthesis, and phenotypic image-based screening.

1.4. Experimental

1.4.1. General Experimental Procedures

Solvents for all chromatography were HPLC grade and used without further purification. Optical rotations were measured on a Jasco P-2000 polarimeter using a 10 mm or 100 mm path length cell at 589 nm. UV Spectra were recorded on a Shimadzu UV-Visible Spectrophotometer (UV-1800) with a path length of 1 cm. NMR spectra were acquired on a Varian Inova 600 MHz spectrometer equipped with a 5 mm HCN triple resonance cryoprobe, and referenced to residual solvent proton and carbon signals (δ\textsubscript{H} 7.26, δ\textsubscript{C} 77.1 for CDCl\textsubscript{3} and δ\textsubscript{H} 3.31, δ\textsubscript{C} 49.0 for CD\textsubscript{3}OD). High-resolution mass spectra were acquired using an Agilent 6230 electrospray ionization (ESI) accurate-mass time-of-flight (TOF) liquid chromatograph-mass spectrometer.

All reactions were performed in oven dried glassware under an inert atmosphere of N\textsubscript{2}. Tetrahydrofuran (THF), was obtained from a Pur-Solv 400 solvent purification system manufactured by Innovative Technology. All reagents were used as purchased without further purification. Thin layer chromatography was performed with Merck Silica gel 60 F254 and visualized with a UV lamp at 254 nm. Crude reaction mixtures were purified using Silica Gel 60 (230 – 400 mesh ASTM).
1.4.2. Fermentation and Isolation

The producing organism, RL10-282-NTS-A, was isolated from a marine sediment sample collected by SCUBA near American Samoa. The strain was originally isolated on NTS medium (20.0 g of agar, 50.0 mg of nalidixic acid, 50.0 mg of cycloheximide, 20.0 g of starch, 0.5g of NaCl, 0.01 g of FeSO₄•7H₂O, 0.5 g MgSO₄•7 H₂O, 0.5 g of K₂PO₄, 1.0 g of KNO₃, 750 mL of 0.2 μm filtered seawater, 250 mL of Milli-Q water). Frozen stocks of environmental isolates were streaked onto fresh Marine Broth plates (37.4 g of Difco Marine Broth, 18 g of agar, 1 L of Milli-Q water) and incubated at 25°C until discrete colonies became visible. Selected colonies were inoculated into 10 mL of modified saline SYP (mSYP) media (10 g starch, 4 g peptone, 2 g yeast extract and 31.2 g instant ocean in 1 L of distilled water). The cultures were stepped up in stages at 7 day intervals by first inoculating 1.5 mL of the 10 mL cell cultures into 50 mL of mSYP (medium-scale), followed by inoculation of 40 mL of these medium-scale cell cultures into 1 L of the same broth also containing 20.0 g of Amberlite XAD-16 adsorbent resin in 2.8 L Fernbach flasks for 7 days. All cultures were incubated at 25°C containing glass beads for 10 mL cultures and stainless steel springs for 50 mL and 1 L cultures and shaken at 200 rpm.

The cells and resin were removed by vacuum filtration using Whatman glass microfiber filters and washed with deionized water. This cell/resin slurry was extracted with 250 mL of 1:1 methanol/dichloromethane, and the organic extract was removed by vacuum filtration and concentrated to dryness in vacuo. The crude organic extract, given extract code RLUS1665, was subjected to solid-phase
extraction (SPE) using a Supelco-Discovery C\textsubscript{18} cartridge (10 g) and eluted using a step gradient of 80 mL of MeOH/H\textsubscript{2}O solvent mixtures (10\% MeOH, 20\% MeOH (A), 40\% MeOH (B), 60\% MeOH (C), 80\% MeOH (D), 100\% MeOH (E) and finally with ethyl acetate (F) to afford seven fractions designated as prefractions A-F. The 10\% MeOH fraction was discarded, and the remaining six were dried in vacuo. The 80\% MeOH prefraction, RLUS1665D, was subjected to RP-HPLC (Phenomenex Synergi Fusion-RP 10 micron, 80 Å, 250 x 4.6 mm, 65:35\% MeOH/H\textsubscript{2}O + 0.02\% formic acid isocratic run over 25 min., 2 mLmin\textsuperscript{-1} flow rate) to afford compounds 3 – 6 which eluted at 5.75, 14.10, 15.70, and 21.95 minutes respectively.

1.4.3. Cytological Profiling and Screening

Cytological profiling was performed at the UC Santa Cruz Chemical Screening Center using our standard screening protocol.\textsuperscript{8} In general, prefractions were tested at two concentrations (1:5 and 1:25 dilutions from master stock solutions). For stock solution preparation protocol see reference 8. Pure compounds were screened as two-fold dilution series (16 dilutions, 100 \( \mu \text{M} \) – 3 nM final testing concentrations). 150 nL of DMSO solutions of test compounds or extracts were added to two separate black-walled clear-bottomed 384 well plates seeded with HeLa cells and incubated at 37\(^\circ\)C for 19 hr. Each plate was fixed and stained with the appropriate stain set (plate 1: Hoechst dye (DNA), anti-phosphohistone H3 antibody (mitotic marker), and EdU (S-phase) plate 2: Hoechst dye (DNA), TMR-phalloidin (actin), anti-tubulin Ab (tubulin); for staining protocol see reference 8) then imaged at four sites per well using an ImageXpress Micro epifluorescent microscope (Molecular
Devices) with a 10x Nikon objective lens. Images were subsequently journaled to extract cell-by-cell features using Molecular Devices MetaXpress software, and these cell-by-cell values converted to final cytological profiles using our in-house data analysis pipeline. Finally, cytological profiles were clustered using Cluster 3.0, and visualized using Java Treeview.

1.4.4. Synthesis of XR344

XR344 was synthesized from diacetyl-2,5-piperazinedione, benzaldehyde and p-anisaldehyde following established literature procedures. The structures of all synthetic products were confirmed by NMR, mass spectrometry and comparison with literature values. Synthetic and naturally occurring XR334 were compared by NMR and LCMS co-injection.

Synthetic scheme for XR334 (3)

(3Z)-1-Acetyl-3-(4-methoxybenzylidene)-2,5-piperazinedione

2,5-Piperazinedione (100 mg, 0.87 mmol) was heated in acetic anhydride (40 mL) at reflux for 4 hours. The crude reaction mixture was concentrated to dryness in vacuo and the product recrystallized from hexane-ethyl acetate to yield 1,4-diacetyl-
2,5-piperazinedione as a tan solid. A solution of 1,4-diacetyl-2,5-piperazinedione (10 mg, 50.5 µmol) in dry THF (1 mL) was cooled to 0°C and a solution of potassium t-butoxide (5.67 mg, 50.5 µmol) and 4-methoxybenzaldehyde (6.88 mg, 50.5 µmol) in t-butanol (1 mL) added dropwise over 30 minutes and allowed to warm to room temperature overnight. The mixture was diluted with ethyl acetate (4mL) and washed with water (2 x 2mL) and saturated brine (2 x 2mL) to afford (3Z)-1-acetyl-3-(4-methoxybenzylidene)-2,5-piperazinedione as a white precipitate. Physical properties and spectra were consistent with published data.

**XR334**

A mixture of (3Z)-1-acetyl-3-(4-methoxybenzylidene)-2,5-piperazinedione (5 mg, 18.25 µmol), caesium carbonate (5.95 mg, 18.25 µmol) and benzaldehyde (1.85 µL, 18.25 µmol) in dimethylformamide (5 mL) was heated at 90°C with stirring in air. After 2 hours the mixture was cooled to room temperature, diluted with ethyl acetate (1 mL) and washed with water (2 x 2mL) and saturated brine (2 x 2mL) to afford crude XR334, which was then subjected to RP-HPLC (Phenomenex Synergi Fusion-RP 10 micron, 80 Å, 250 x 4.6 mm, 65:35% MeOH/H₂O + 0.02% formic acid isocratic run over 10 min., 2 mLmin⁻¹ flow rate) to afford XR334 as a white solid with a retention time of 5.75 minutes. Physical properties and spectra were consistent with published data.
1.4.5. Co-injection of synthetic and natural product XR334 (3)
1.4.6. $^1$H NMR (600MHz, CDCl$_3$) spectrum of synthetic and isolated XR334
1.4.7. $^{13}$C NMR (150 MHz, CD$_3$OD) spectrum of XR334 (3)
1.4.8. $^1$H NMR (600MHz, CD$_3$OD) spectrum of nocapyrone B (4)
1.4.9. \( ^1 \)H NMR (600MHz, CD\(_3\)OD) spectrum of nocapyrone H (5)
1.4.10. $^1$H NMR (600MHz, CD$_3$OD) spectrum of nocapyrone L (6)
1.4.11. Expanded CP fingerprint of nocaprynone L (6)
1.4.12. HPLC trace for reverse phase stage fractionation of RLUS1665D
1.4.13. Dilution Series of XR334 and Nocapyrone L
1.5. References


(19) Hayashi, Y.; Takeno, H.; Chinen, T.; Muguruma, K.; Okuyama, K.; Taguchi,

(20) *AntiMarin Database*; Blunt, J. W.; Munro, M. H. G.; Laatsch, H., Eds.; University of Canterbury: Christchurch, New Zealand, and University of Göttingen; Göttingen, Germany, 2006.


Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; et


Chapter 2

The Marine Mammal Microbiome Yields a Novel Antibiotic with Potent Activity Against *Clostridium difficile*
Table of Contents

2. The Marine Mammal Microbiome Yields a Novel Antibiotic with Potent Activity Against *Clostridium difficile* ......................................................... 46
   2.1. Introduction ........................................................................................................ 46
   2.2. Results and Discussion ....................................................................................... 49
       2.2.1. Sample Collection and Isolate Prioritization for Novel Bioactives Discovery....... 49
       2.2.2. Biological Screening of Crude Extracts. ....................................................... 51
       2.2.3. Structure Elucidation. ................................................................................. 53
       2.2.4. Biological Activity....................................................................................... 56
   2.3. Conclusion .......................................................................................................... 61
   2.4. Experimental ....................................................................................................... 63
       2.4.1. General Experimental Procedures ................................................................. 63
       2.4.2. Cultivation of Bacteria................................................................................... 63
       2.4.3. DNA Isolation, PCR Amplification and Sequencing. .................................... 64
       2.4.4. Phylogenetic Analysis. ................................................................................. 65
       2.4.5. Molecular Networking Experimental: ......................................................... 69
       2.4.6. Extraction of Cultivated Isolates. ................................................................. 70
       2.4.7. Fermentation and Isolation. ......................................................................... 71
       2.4.8. Flow Cytometry. ....................................................................................... 72
       2.4.9. Bacterial Strains. ......................................................................................... 73
       2.4.10. High-Throughput Antibacterial Inhibition Assay.......................................... 73
       2.4.11. Cytological Profile Screening and Image Analysis......................................... 75
2.4.12. Extended Structure Elucidation ................................................. 77
2.4.13. Configurational Analysis ................................................................. 82
2.4.14. NMR Chemical Shifts ................................................................. 87
2.4.15. NMR Spectra .............................................................................. 88
2.5. References ..................................................................................... 99
2. The Marine Mammal Microbiome Yields a Novel Antibiotic with Potent Activity Against *Clostridium difficile*

2.1. Introduction

Historically, Actinobacteria have been a prolific source of diverse secondary metabolites and have provided the inspiration for numerous clinically approved therapeutics.\(^1\) Many marine-derived Actinobacteria have been isolated from invertebrates or found in association with other marine organisms.\(^2-5\) These organisms have been proposed to act as symbionts, protecting the host from pathogenic invasion through the production of bioactive small molecules.\(^2,6\) In a systematic survey of the biosynthetic gene capacity of prokaryotes, Actinobacteria were demonstrated to possess the highest mean number of biosynthetic gene clusters (BGCs) per strain.\(^7,8\) These results reinforce previous empirical observations, which have long highlighted Actinobacteria as the single most important bacterial phylum for small molecule drug discovery.\(^9-11\)

The human microbiome has unique potential for the discovery of bioactive metabolites because the BGCs present in bacterial isolates from humans differ significantly from those found in isolates from non-humans.\(^2-5\) Production of small molecules by commensal bacteria mediates microbe-microbe and microbe-host interactions and has been documented in a diverse array of host systems.\(^2,12-14\) Inspired by the chemical diversity present in microorganisms isolated from these diverse environments, we examined previously the chemical production of Actinobacteria isolated from fish intestines, leading to the discovery of a novel
antimicrobial agent and a biofilm inhibitor. We hypothesized that extension of this analysis to marine mammals would provide access to host-associated bacteria with unique BGC capacities compared with either free-living environmental bacteria, or bacteria from the microbiota of terrestrial hosts.

Aquatic environments provide a unique opportunity for the discovery of novel chemical scaffolds because temperature fluctuations, nutrient concentrations, and the dispersion of chemicals differ significantly compared with terrestrial environments. The specific heat capacity of water is 25 times higher than air, so temperature changes are much more gradual in aquatic environments and typically experience less variability in a given location. The role of temperature and light have been implicated in taxon diversity; in temperate oceans, shorter days such as those during the winter season typically have higher microbial diversity. By comparison, terrestrial diversity exhibits milder shifts in diversity with seasonal variation. As with heat, oxygen is also a limited resource in aquatic environments, resulting in increased bacterial nitrate respiration. Salinity has also been a driving force for microbial diversity, enriching or excluding particular taxa. Phylogenetic analysis of 111 geographically and environmentally diverse samples revealed that salinity was the distinguishing factor for community structure. A variety of abiotic factors have worked to shape differentially both organisms that inhabit aquatic and terrestrial environments and their microbiota.

Habitat also has a general impact on community composition; among 151 animals, marine hosts exhibited significantly richer microbial diversities than their
closest terrestrial relatives. Exposure to diverse environmental factors has been associated with increased intestinal microbiota diversity, and reduced exposure to environmental microbes is a proposed mechanism for the loss of microbial diversity in captive animals. This connection is exemplified by the observation that marine mammals with extensive global migration patterns are exposed to drastic changes in the environment and also possess increased diversity in microbial community structure compared with non-migratory species. Interestingly, 70% of the near full-length sequences obtained from cetaceans are significantly distinct by 16S rRNA sequencing to published sequences in the NCBI database, indicating that these marine mammals are sources of novel species and possibly small-molecule production. Moreover, longitudinal sampling from the mouth down the gastrointestinal tract of marine dolphins and sea lions revealed increasing differences between the gastrointestinal bacterial community composition compared with the surrounding planktonic community. Thus, marine mammals clearly possess complex and understudied microbial communities.

To date, most marine mammal microbiome studies have focused on the detection and analysis of pathogens, rather than characterization of their commensal bacteria. These surveys have identified a wide array of bacterial pathogens that are also important in human disease, such as *Pseudomonas aeruginosa* and *Staphylococcus aureus* (including MRSA). Research focusing on the human microbiome has amassed a wealth of knowledge, implicating its involvement with overall host health. The individual microbiome is as unique as a fingerprint and
hides clues their ancestry and daily lives. This research has led to meta-analyses of marine mammal microbiomes, illuminating a previously unseen aspect of their secretive lives. Marine mammals are sentinel creatures that live at the interface of the aquatic and terrestrial worlds and possess distinct commensal bacteria. Considering the BGC capacity of host-associated microbes and the exposure of marine mammals to distinctive abiotic and biotic factors, we hypothesized that wild marine mammal microbiomes would be a unique source for antimicrobial discovery.

2.2. Results and Discussion

2.2.1. Sample Collection and Isolate Prioritization for Novel Bioactives Discovery.

Samples were collected from various marine mammals, isolated, extracted and prioritized to optimize the identification of small molecule antibiotics (Figure 2.1). In collaboration with the Long Marine Laboratory Marine Mammal Stranding Network, samples of intestinal contents from the carcasses of five marine mammals (harbor porpoise (MMA), harbor seal (MMB), California sea lion (MMC), infant harbor porpoise (MMD), and infant dolphin (MME)) were sampled for their culturable microbiota (Table 2). From the MMA samples, 48 isolates were obtained using lean isolation media designed to enhance for Actinobacteria. Of these, 28 were selected for further sequencing analysis. 16S rRNA sequencing showed that the majority (22/28) of the sequenced bacteria were Firmicutes, while Actinobacteria comprised 21.4% (Table 3). The lowest sequence identity observed in comparison with published
strains was 94.4% with a *Bacillus oceanisediminis* strain. To prioritize organisms for chemical evaluation, strains with low sequence identity were selected, as well as strains from genera considered rare (understudied) in terms of Actinobacterial natural products production, such as *Micromonospora*.

Metabolites from colonies from agar plates were extracted with organic solvents and analyzed by direct infusion tandem mass spectrometry. A molecular network was assembled based on the fragmentation patterns of the extracted metabolites produced from those colonies, using Global Natural Products Social Sample Collection, Extract Preparation & Screening, Structure Elucidation, Pure Compound Assay.

Figure 2.1 Overview of project workflow. Isolates collected from marine mammal intestinal contents were fermented, and fractionated into 384 well plates for bioassays. Phocenamicin was extracted from organism FI1036 and analyzed as a pure compound.
molecular networking (GNPS), (Experimental Figure 2.8). Metabolites were characterized as unique if they were produced by only a single organism. The five organisms within the sample set that displayed the highest number of unique metabolites were also prioritized. In total, eleven organisms were fermented at the 1L scale, extracted, and fractionated for biological screening.

### 2.2.2. Biological Screening of Crude Extracts.

Image based screening of the crude extracts against our biofilm inhibition assay in *Vibrio cholera*, suggested that extracts RLF1039D, RL1041E and RLF1042F possessed biofilm inhibition properties. High-throughput screening against a panel of 15 BLS1 and BSL2 pathogens was performed using our in-house BioMAP assay, which clusters antibiotic compounds through normalizing their relative MIC values for a concentration independent analysis. This indicated that seven of the eleven prioritized organisms possessed antimicrobial activity against one or more bacterial strains. This hit rate is comparable to our traditional library of non-marine mammal-associated bacteria, for which 55% of all tested strains show some degree of antibacterial activity against one or more test strains. Secondary screening of dilution series of active fractions (16×two-fold dilutions) highlighted fraction RLF1036F from *Micromonospora auratinigra* as a high potency lead, with activity against Gram-positive bacteria down to the lowest tested concentration (Figure 2.2). This fraction was therefore prioritized for further analysis utilizing our standard natural product isolation protocol. After two stages of fractionation, a single bioactive component was isolated possessing mass spectrometric m/z features of 1071.4 and
1093.4, consistent with the [M+H]$^+$ and [M+Na]$^+$ adducts for a molecule with a mass of 1070.4 Da.

<table>
<thead>
<tr>
<th>NP Extract</th>
<th>S. aureus</th>
<th>MRSA</th>
<th>S. Subtilis</th>
<th>E. faecium</th>
<th>L. ivanovii</th>
<th>S. epidermidis</th>
<th>A. baumanii</th>
<th>P. aeruginosa</th>
<th>V. cholerae</th>
<th>E. coli</th>
<th>P. aerogenes</th>
<th>E. coli</th>
<th>S. typhimidium</th>
<th>Y. pseudotuberculosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLFI_1034F</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1035A</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1035D</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1035E</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1036E</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1036F</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1036G</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1037B</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1037F</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1040E</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1042B</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1042C</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1042D</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1042E</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>RLFI_1044B</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
</tbody>
</table>

**Figure 2.2** Crude extract BioMAP dilution profiles. Natural product fractions on Y axis. Numbering denotes organism, RLFI denotes origin of organism, letters A-G denotes polarity of solvent used for extraction.
2.2.3. Structure Elucidation.

Evaluation of the high-resolution mass spectrometry (HRMS) m/z value for the precursor ion, coupled with examination of the isotopic distribution, suggested a molecular formula of C$_{56}$H$_{75}$ClO$_{18}$. The planar structure was assembled by NMR analysis, using a combination of $^1$H, $^{13}$C, gCOSY, gHSQC, gHMBC, and TOCSY spectra (Figure 2.3A). These analyses revealed a novel spirotetronate macrocyclic core with a carbon skeleton containing 23 chiral centers that was named phocoenaminic (Figure 2.4); (for full structure determination description, see extended structure elucidation in experimental).

![Figure 2.3](image.png)

Figure 2.3 Structure elucidation of phocoenaminic. (A) Key NMR correlations used in the planar structure elucidation. (B) Newman projection along C$_{21}$(front) and C$_{31}$ and key coupling constant values to determine bond angles. (C) Newman projection along C$_{31}$(front) and C$_{32}$ and key coupling constant values. (D) Newman projection along C$_{33}$ and C$_{34}$ and NOE correlations.
Extensive exploration of crystallization conditions for single crystal X-ray diffraction proved unsuccessful, as has been observed for a number of other spirotetronate natural products. However, determination of the absolute configuration for all 23 configurational centers in phocoenamicin was accomplished using a combination of spectroscopic and chemical derivatization methods. Detailed discussion for the complete assignment is presented in the extended structure elucidation section. In brief, the constitution and relative configuration of the polyketide core of the molecule were determined using NMR methods, while the carbohydrate monomers were determined using a combination of NMR analyses and degradation and derivatization studies in comparison with commercially available standards. The most challenging portion of structure elucidation was assignment of the configurations of the chiral centers on the diol sidechain (C31 – C35). Extensive signal overlap in the proton NMR spectrum precluded the facile extraction of coupling constants for this key region. Application of a selective 1D TOCSY experiment irradiating the proton at 2.45 ppm (C21) on the bridgehead of the alkyl chain, in conjunction with data derived from a DQF-COSY spectrum in this region, provided clear evidence for the coupling constants associated with protons at 1.98 and 1.75 (C31), 3.87 (C32), and 2.30 and 1.85 (C22).

For the methylene at C31, a large coupling constant of 9.6 Hz was observed between the proton at 1.98 ppm and the adjacent proton at 2.45 ppm on C21, suggesting a dihedral angle of close to 0° or 180°. A small coupling constant of 3.1 Hz between the proton at 2.45 ppm on C21 and 1.75 ppm on C31 indicated a dihedral angle
angle of ~60°. Taken together, these coupling constants suggested the atom arrangement depicted in Fig. 2.3B. This was further confirmed by the presence of reciprocal ROE correlations between the resonances at 2.45 and 1.75 ppm. Extending along the sidechain, the proton at 1.98 ppm on C31 possessed a medium-sized coupling constant (4.8 Hz) to the adjacent proton on C32 at 3.87 ppm, with corresponding ROE correlations indicating a small bond angle of ~60°.

Additionally, a large coupling constant between the protons at 3.87 ppm and 1.75 ppm indicated a large bond angle of approximately 180° (Figure 2.3C). With these bond angles established, all four diastereomers were built and modeled in Discovery Studio 4.0. Diastereomers exhibiting S stereochemistry at C33 or R stereochemistry at C32 exhibited conformers that sacrificed at least one observed ROE correlation in each conformation (Experimental Figure 2.12). Only one

![Chemical structures](image)

Figure 2.4 Structure of phocoenaminic, closest analogue maklamicin, and fidaxomicin.
diastereomer (32S, 33R) aligned with all the observed ROE correlations, thus completing the bond orientation at C32 (Figure 2.3D) and the full relative configurational analysis of the polyketide portion (Figure 2.4)

2.2.4. Biological Activity.

Antibacterial screening of phocoenamicin against a panel of clinically-relevant bacteria revealed potent activity against our panel of Gram-positive bacteria (Table 1). Comparison of phocoenamicin to other structures within the spirotetronate class revealed that while phocoenamicin shares a similar macrolide backbone to several other spirotetronates including maklamicin (Figure 2.4), it possesses several unique structural features including the diol sidechain and the presence of two 6-deoxy sugars not previously observed within this class. Phocoenamicin also shares several key features with fidoxamicin (Figure 2.4), a promising narrow spectrum antibiotic for treating infections of Clostridium difficile, an antibiotic-associated

<table>
<thead>
<tr>
<th>Organism</th>
<th>Phoc. (µM)</th>
<th>Vanc. (µM)</th>
<th>Rif. (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>&gt;66</td>
<td>&gt;66</td>
<td>8.3</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>&gt;66</td>
<td>&gt;66</td>
<td>50</td>
</tr>
<tr>
<td>B. subtilis</td>
<td>2.8</td>
<td>0.15</td>
<td>0.12</td>
</tr>
<tr>
<td>E. faecium</td>
<td>11</td>
<td>3.6</td>
<td>4.47</td>
</tr>
<tr>
<td>S. aureus (MSSA)</td>
<td>1.0</td>
<td>0.69</td>
<td>0.010</td>
</tr>
<tr>
<td>S. aureus (MRSA)</td>
<td>1.0</td>
<td>0.87</td>
<td>0.0081</td>
</tr>
<tr>
<td>S. epidermidis</td>
<td>19</td>
<td>20</td>
<td>0.020</td>
</tr>
<tr>
<td>P. alcalifaciens</td>
<td>&gt;66</td>
<td>&gt;66</td>
<td>17</td>
</tr>
<tr>
<td>O. anthropi</td>
<td>&gt;67</td>
<td>&gt;66</td>
<td>22</td>
</tr>
<tr>
<td>E. aerogenes</td>
<td>&gt;68</td>
<td>&gt;66</td>
<td>17</td>
</tr>
<tr>
<td>A. baumanii</td>
<td>&gt;69</td>
<td>&gt;66</td>
<td>6.9</td>
</tr>
<tr>
<td>V. cholerae</td>
<td>&gt;70</td>
<td>&gt;66</td>
<td>28</td>
</tr>
<tr>
<td>S. typhimurium</td>
<td>&gt;71</td>
<td>&gt;66</td>
<td>33</td>
</tr>
<tr>
<td>Y. pseudotuberculosis</td>
<td>&gt;72</td>
<td>&gt;66</td>
<td>8.3</td>
</tr>
<tr>
<td>C. difficile</td>
<td>2.6</td>
<td>2.9</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 1 Pure compound average MIC values. All assays run in triplicate, abbreviations used: phocoenamicin (phoc.), vancomycin (vanc.), and rifampicin (rif.).
pathogen that causes ~500,000 infections and is linked to >29,000 deaths per year.\textsuperscript{33} Both compounds possess a 6-deoxy sugar with an acetyl linkage to a chlorinated phenol, prompting reevaluation of phocoenamicin against \textit{C. difficile}. \textit{Clostridium} species are relevant to marine mammals and have been isolated from intestinal and fecal samples from cetaceans, including dolphins and whales.\textsuperscript{13,34} Limited information is available on mortality rates of marine mammals from \textit{C. difficile} infections. Although an emerging threat to humans, this pathogen has limited treatment options with current therapy consisting of an aggressive dosage of vancomycin and/or fecal transplants to try to re-establish microbiome health.\textsuperscript{35} Screening of phocoenamicin against \textit{C. difficile} (ATCC 700057) according to CSLI guidelines revealed comparable activity (MIC: 2.60 \(\mu\)M) to vancomycin (MIC: 2.87 \(\mu\)M),\textsuperscript{36} making this an encouraging lead compound for further development.

To examine the effect of phocoenamicin on mammalian cell development, an image-based cytological profiling platform was applied.\textsuperscript{37,38} Compound-treated HeLa cells remained viable after 24 h of incubation with phocoenamicin, with only the two highest concentrations (73 \(\mu\)M and 146 \(\mu\)M) showing subtle changes in cell morphology (Figure 2.5B). At these concentrations, phocoenamicin clustered with known ionophores and mitochondrial decouplers, suggesting that the antibacterial mode of action (MOA) is related to modification of ionic gradients across the cell membrane (Figure 2.5C). To determine MOA of phocoenamicin, a \textit{Bacillus subtilis} essential gene knockdown (CRISPRi) library and transposon insertion library was screened against phocoenamicin (Figure 2.5D). Previous studies have shown these
two methods as platforms for discovery of drug target protein.\textsuperscript{39,40} As expected, strains over-producing multidrug efflux pump (Mdt) were identified as phocoenamicin-resistant (up to 4 X MIC) ones. However, a direct target of phocoenamicin from either CRISPRi screen or transposon insertion library screen could not be found (Figure 2.5E). Thus, these results also support the idea that the target of phocoenamicin might be a non-protein essential cell function, likely ionic gradients across the cell membrane suggested from cytological profiling.
Figure 2.5 Cytological profiling of phocoenamicin. (A) Complete heatmap of 5,000 compounds (Y-axis) hierarchically clustered and evaluated against 248 features (X-axis). Positive deviations to DMSO controls represented in yellow, negative deviations represented in blue. (B) Fluorescence images of drug-treated wells. Tubulin is represented in green, mitotic cells in cyan, actin in red, and DNA in blue. (C) Expanded fingerprints of phocoenamicin and neighboring cluster. Genetic screen to identify cellular target of phocoenamicin. (D) Screen of essential gene knockdown library in Bacillus subtilis against phocoenamicin. Relative fitness distribution of CRISPRi essential gene knockdown strains grown on plates containing sub-inhibitory concentrations of trimethoprim (left) and phocoenamicin (right). Screens were performed and relative fitness determined using previous methods. Strains sensitized to the antibiotic by depletion of the essential gene target are expected to have a lower relative fitness. The target of trimethoprim was accurately identified as DfrA, however no apparent essential gene target was observed for phocoenamicin. (E) Identification of transposon insertion sites of the strains showing resistance to phocoenamicin. B. subtilis transposon insertion library was plated on LB agar plate containing 4 X MIC of phocoenamicin. After 16 hr incubation, 95 phocoenamicin-resistant colonies were isolated and transposon insertion site was mapped for 20 strains. All 20 transposon insertions were located upstream of mdrR-mdtP operon which induce overexpression of mdtP, multidrug efflux pump gene. Inserted positions (red line) were labeled with coordinates and isolate number.
In other ionophores, varying calcium or magnesium concentration, or altering the pH of the growth medium can have drastic effects on the biological activity.\textsuperscript{41} However, when these variables were systematically changed in the growth medium, no significant effect was observed on the MIC of phocoenamicin against \textit{S. aureus} (Table 4). Thus, calcium does not induce conformational changes that affect its bioactivity as proposed for daptomycin, a known calcium dependent antimicrobial that is active against Gram-positive bacteria.\textsuperscript{42} Gram-positive bacteria such as \textit{S. aureus} and \textit{C. difficile} can survive in acidic environments through a variety of mechanisms such as changes to membrane composition and proton pumps.\textsuperscript{43} These defense mechanisms can also make them resistant to antibiotics; nonetheless, these results indicate that phocoenamicin retains its chemical stability and potency in acidic conditions against \textit{S. aureus}. Considering both the cytological profiling and bioMAP profiles, fluorescence-activated cell sorting (FACS) was performed to evaluate phocoenamicin as an ionophore. Phocoenamicin was determined to have the ability to effectively depolarize the membrane when compared to the positive control carbonyl cyanide \textit{m}-chlorophenyl hydrazine, without forming pores in the cell membrane (Figure 2.6).\textsuperscript{44}
Figure 2.6 Membrane polarity and permeability assays. (A) Microscopy images of *S. aureus* cells treated with 70% ethanol, ½ MIC of phocoenaminic or DMSO and membrane permeable (DRAQ5) and impermeable dye (PI) show the compromised integrity of the ethanol treated cells and the intact membranes of compound and DMSO treated cells. (B) Flow cytometry results using DiCO$_2$ shows that phocoenaminic polarizes the membrane compared to the negative control (DMSO), and the positive control (CCCP).

2.3. Conclusion

Unique access to the Marine Mammal Stranding Network allowed the exploration of several marine mammal microbiotas. Isolation of 48 culturable microorganisms from an otherwise healthy harbor porpoise prior to stranding, combined with biological assays, showed the antimicrobial potential of microbiota-based collections. Limited information is available about the compounds produced by mammalian bacterial isolates, or the likely endogenous roles of those compounds. Prioritization of *Micromonospora auratinigra* from the harbor porpoise microbiota revealed the presence of a new glycosylated polyketide antibiotic, phocoenaminic, with potent activity against a broad panel of Gram-positive organisms including the intestinal pathogen *C. difficile*. The discovery of phocoenaminic has highlighted the
potential of marine mammal commensal microbes for drug discovery and host
protection.

Phocoenamicin possesses comparable activity to vancomycin, the current
frontline antibiotic against *C. difficile* infections. Interrogation of the MOA for
phocoenamicin using a suite of approaches demonstrated that phocoenamicin
functions by disrupting membrane potential without affecting membrane integrity.
Vancomycin has previously been shown to have no impact on membrane potential,\(^{45}\)
suggesting disparate MOAs for these two compounds. A second emerging treatment
for *C. difficile*, fidaxomicin, shares a number of similar structural features with
phocoenamicin. Although both phocoenamicin and fidaxomicin have stable MICs
across varying cation and pH conditions, lack of sensitivity of an RNA polymerase
knock-down strain in *B. subtilis* suggests that phocoenamicin possesses a different
mechanism of action to fidaxomicin, which has been shown to target RNA
polymerase.\(^{46}\) Together these results suggest phocoenamicin functions by an
alternative mechanism to any of the existing treatments for *C. difficile* used in clinical
practice. The discovery of this novel spirotetronate antibiotic demonstrates the value
in exploring niche microbiota environments for novel compound discovery, and
provides incentive for further investigation of the understudied host-microbe
environment of marine mammals.
2.4. Experimental

2.4.1. General Experimental Procedures.

Chromatography solvents were HPLC-grade, and were used without further purification. Optical rotations were measured on a Jasco P-2000 polarimeter using a 10-mm path length cell at 589 nm. UV spectra were recorded on a Shimadzu UV-visible spectrophotometer (UV-1800) with a path length of 1 cm. NMR spectra were acquired on a Varian Inova 600 MHz spectrometer equipped with a 5 mm H/C/N triple resonance cryoprobe and referenced to residual solvent proton and carbon signals ($\delta_H$ 2.05, $\delta_C$ 29.84, for acetone-d6 and $\delta_H$ 7.26, $\delta_C$ 77.16 for chloroform-d). HRMS data were acquired using an Agilent 6230 electrospray ionization accurate-mass time-of-flight liquid chromatograph mass spectrometer.

2.4.2. Cultivation of Bacteria.

Marine mammal intestinal contents were transferred to sterile Falcon tubes, 1 mL of sterile Milli-Q water was added, and the samples were vortexed for 1 min. Five marine mammals were analyzed (see table below). Four solid agar media were used for microbial isolation: actinomycete isolation agar (Difco), SNS, and modified NTS and HVS$^{16}$. All isolation plates were prepared with sterile sea water and supplemented with 50 mg/L of both cyclohexamide and nalidixic acid. Intestinal contents were plated using three different methods: 1) the mixture was serially stamped onto solid agar with a sterile swab, 2) the mixture was diluted with 1 mL of sterile Milli-Q water and 100 $\mu$L of the resulting mixture was spread onto the plate.
surface, and 3) the mixture was diluted with 10 mL of sterile Milli-Q water and 100 µL of the resulting mixture was spread onto the plate surface. Cultures were incubated at room temperature and bacterial colonies displaying desired morphologies were sub-cultured on Difco Marine Broth solid agar plates until pure. Typical incubation times for the appearance of colonies from isolation plates ranged from 30–90 days.

Table 2 Marine mammal necroscopy details. Whole dead intact mammals were recovered as part of the California Marine Mammal Stranding Network, and stored at -20°C until dissection. All instruments were flame sterilized prior to dissection, instruments were flame sterilized after each cut. The stomach and gastrointestinal tract were separated from the body cavity. 1mL aliquots sampled through the digestive tract were transferred to sterile 10 mL Falcon tubes at discrete portions of the gut: the posterior portion of the intestine, the mid-intestine, the stomach, and the duodenum, as appropriate for each specimen.

<table>
<thead>
<tr>
<th>Code</th>
<th>Specimen</th>
<th>Species</th>
<th>Common Name</th>
<th>details</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMA</td>
<td>LMLPP2011SEP29</td>
<td>Phocoena phocoena</td>
<td>Harbor porpoise</td>
<td>Adult female, killed by dolphins</td>
</tr>
<tr>
<td>MMB</td>
<td>LMLPV2013MAR03</td>
<td>Phoca vitulina</td>
<td>Harbor Seal</td>
<td>Adult female, possible infection</td>
</tr>
<tr>
<td>MMC</td>
<td>LMLZC2013APR10</td>
<td>Zalophus californianus</td>
<td>California Sea Lion</td>
<td>Adult female, emaciated</td>
</tr>
<tr>
<td>MMD</td>
<td>LMLPP2013JUL23</td>
<td>Phocoena phocoena</td>
<td>Harbor porpoise</td>
<td>Infant, killed by dolphins</td>
</tr>
<tr>
<td>MME</td>
<td>LMLDC2013JUL31</td>
<td>Dolphinus capensis</td>
<td>Common dolphin</td>
<td>Infant, death by separation</td>
</tr>
</tbody>
</table>

2.4.3. DNA Isolation, PCR Amplification and Sequencing.

Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega) by picking a single colony of the cultured strains according to manufacturer's instructions. For PCR amplification of 16S rRNA gene, the primer
pair 8F (5’-AGAGTTTGATCTGGCTCAG-3’) and 1492R (5’-GGTTACCTGTACGACTT-3’) was used. Platinum® Taq DNA Polymerase, High Fidelity (Invitrogen) was used for amplification according to manufacturing protocol Pub. No. MAN0000948 Rev. A.0 to achieve a total reaction volume of 25 µL. PCR was performed on an Eppendorf Mastercycler Personal thermocycler under the following conditions: initial denaturation 95 °C for 5 min, 35 cycles of denaturation at 95 °C for 1 min, annealing at 50 °C for 1 min, extension at 72 °C for 90 s, and a final extension at 72 °C for 10 min. After confirmation by gel electrophoresis (1% agarose gel in 1X TAE buffer), the PCR products were purified with QIAQuick PCR Purification Kit (Qiagen) and sent directly to Sequetech Corporation for sequencing using the same PCR primers described above plus an additional middle primer 341F (5’-CCTACGGGAGGCAGCAG-3’). DNA sequences have been deposited to GenBank with accession numbers KY580789 - KY580816.

2.4.4. Phylogenetic Analysis.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The bootstrap consensus tree inferred from 2000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) are shown next to the branches (Figure 2.7). Initial tree(s) for the heuristic search were obtained automatically by applying
Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The analysis involved 44 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There was a total of 1195 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.\textsuperscript{50}
<table>
<thead>
<tr>
<th>Organism Name</th>
<th>Strain</th>
<th>Accession Number</th>
<th>Closest NCBI Strain</th>
<th>% identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bacillus algicola</em></td>
<td>MMA10XX</td>
<td>KY580815</td>
<td>NR_029077.1</td>
<td>98.6</td>
</tr>
<tr>
<td><em>Bacillus firmus</em></td>
<td>MMA1002</td>
<td>KY580790</td>
<td>NR_025842.1</td>
<td>99.9</td>
</tr>
<tr>
<td><em>Bacillus soli</em></td>
<td>MMA1006</td>
<td>KY580791</td>
<td>NR_025591.1</td>
<td>98.6</td>
</tr>
<tr>
<td><em>Streptomyces coelicolor</em></td>
<td>MMA1009</td>
<td>KY580792</td>
<td>NR_116633.1</td>
<td>100</td>
</tr>
<tr>
<td><em>Bacillus hwajinpoensis</em></td>
<td>MMA1010</td>
<td>KY580793</td>
<td>NR_025264.1</td>
<td>100</td>
</tr>
<tr>
<td><em>Bacillus oceanisediminis</em></td>
<td>MMA1012</td>
<td>KY580794</td>
<td>NR_117285.1</td>
<td>99.6</td>
</tr>
<tr>
<td><em>Bacillus atrophaeus</em></td>
<td>MMA1013</td>
<td>KY580796</td>
<td>NR_024689.1</td>
<td>99.8</td>
</tr>
<tr>
<td><em>Bacillus atrophaeus</em></td>
<td>MMA1014</td>
<td>KY580797</td>
<td>NR_024689.1</td>
<td>99.9</td>
</tr>
<tr>
<td><em>Micromonospora zamorensis</em></td>
<td>MMA1017</td>
<td>KY580798</td>
<td>NR_108479.1</td>
<td>99.5</td>
</tr>
<tr>
<td><em>Mycobacterium gadium</em></td>
<td>MMA1019</td>
<td>KY580799</td>
<td>NR_026087.1</td>
<td>99.3</td>
</tr>
<tr>
<td><em>Bacillus megaterium</em></td>
<td>MMA1020</td>
<td>KY580800</td>
<td>NR_043401.1</td>
<td>99.6</td>
</tr>
<tr>
<td><em>Bacillus megaterium</em></td>
<td>MMA1021</td>
<td>KY580801</td>
<td>NR_117473.1</td>
<td>99.9</td>
</tr>
<tr>
<td><em>Bacillus oceanisediminis</em></td>
<td>MMA1023</td>
<td>KY580802</td>
<td>NR_117285.1</td>
<td>99.6</td>
</tr>
<tr>
<td><em>Bacillus oceanisediminis</em></td>
<td>MMA1024</td>
<td>KY580803</td>
<td>NR_117285.1</td>
<td>98.4</td>
</tr>
<tr>
<td><em>Bacillus oceanisediminis</em></td>
<td>MMA1025</td>
<td>KY580804</td>
<td>NR_117285.1</td>
<td>94.4</td>
</tr>
<tr>
<td><em>Fictibacillus phosphorivorans</em></td>
<td>MMA1027</td>
<td>KY580805</td>
<td>NR_118455.1</td>
<td>99.5</td>
</tr>
<tr>
<td><em>Streptomyces exfoliatus</em></td>
<td>MMA1029</td>
<td>KY580806</td>
<td>NR_041229.1</td>
<td>100</td>
</tr>
<tr>
<td><em>Bacillus atrophaeus</em></td>
<td>MMA1033</td>
<td>KY580807</td>
<td>NR_07016.1</td>
<td>99.8</td>
</tr>
<tr>
<td><em>Paenibacillus glucanolyticus</em></td>
<td>MMA1035</td>
<td>KY580808</td>
<td>NR_113748.1</td>
<td>99.7</td>
</tr>
<tr>
<td><em>Micromonospora coxensis</em></td>
<td>MMA1036</td>
<td>KY580809</td>
<td>NR_041350.1</td>
<td>98.8</td>
</tr>
<tr>
<td><em>Micromonospora auratinigra</em></td>
<td>MMA1037</td>
<td>KY580810</td>
<td>NR_028659.1</td>
<td>99</td>
</tr>
<tr>
<td><em>Paenibacillus glucanolyticus</em></td>
<td>MMA1039</td>
<td>KY580811</td>
<td>NR_113748.1</td>
<td>99.8</td>
</tr>
<tr>
<td><em>Bacillus atrophaeus</em></td>
<td>MMA1041</td>
<td>KY580812</td>
<td>NR_075016.1</td>
<td>99.9</td>
</tr>
<tr>
<td><em>Bacillus atrophaeus</em></td>
<td>MMA1042</td>
<td>KY580813</td>
<td>NR_075016.1</td>
<td>99.9</td>
</tr>
<tr>
<td><em>Bacillus atrophaeus</em></td>
<td>MMA1043</td>
<td>KY580814</td>
<td>NR_075016.1</td>
<td>99.9</td>
</tr>
<tr>
<td><em>Bacillus atrophaeus</em></td>
<td>MMA1044</td>
<td>KY580815</td>
<td>NR_075016.1</td>
<td>99.9</td>
</tr>
<tr>
<td><em>Bacillus atrophaeus</em></td>
<td>MMA1045</td>
<td>KY580816</td>
<td>NR_075016.1</td>
<td>99.9</td>
</tr>
</tbody>
</table>
Figure 2.7 Molecular phylogenetic analysis by maximum likelihood method.
2.4.5. Molecular Networking Experimental:

Bacteria were extracted then directly infused into the mass spectrometer using a Triversa nanomate-electrospray ionization source (Advion Biosystems) coupled to a 6.42 T Thermo LTQ-FT-ICR mass spectrometer. Briefly, bacterial colony plugs were extracted with 100 µL of 50:50 MeOH:H2O and 100 µL of butanol for 1 hr at rt. Following this extraction, samples were spun down at 10,000 rpm for 2 min and 10 µL was diluted into 1 mL of 50:50 ACN:H2O. This solution was spun down at 10,000 rpm for 2 min and dilute another 10 fold and then directly infused using a back pressure of 0.35-0.5 psi and a spray voltage of 1.3-1.45 kV. FT-MS and ion trap MS/MS spectra were acquired using Tune Plus software version 1.0 and Xcalibur software version 1.4 SR1. The instrument was tuned on m/z 816, the 15+ charge state of cytochrome C. The instrument scan cycle consisted of one 10 min segment, during which a profile FT scan with a resolution of 25,000 was cycled with four data-dependent scans in the ion trap. The data-dependent scan iteratively cycled through the top four most intense ions from the FT scan, after which they were placed on an exclusion list for 600 s. Data was converted using MSConvert (Proteowizard) to 32-bit mzXML with peak picking level 1.

A molecular network was created using the online workflow at GNPS. The data was then clustered with MS-Cluster with a parent mass tolerance of 1.0 Da and a MS/MS fragment ion tolerance of 0.3 Da to create consensus spectra. A network was then created where edges were filtered to have a cosine score above 0.5 and more than 6 matched peaks. Further edges between two nodes were kept in the network if
and only if each of the nodes appeared in each other's respective top 10 most similar nodes.

Figure 2.8 Molecular networking of bacterial extracts. Colored nodes are metabolites produced only by one organism, nodes represented in grey are metabolites shared by many organisms.

2.4.6. Extraction of Cultivated Isolates.

Purified bacterial colonies were grown in 1 L of modified SYP broth (1 L Milli-Q water, 32.1 g Instant Ocean™, 10 g starch, 4 g peptone, 2 g yeast extract) with 20 g of Amberlite XAD-16 resin for 10 days at 27 °C. Culture broth and resin slurries were filtered through glass microfiber filters, washed with water (3×200 mL) and the cells, resin, and filter paper were extracted with 1:1 methanol/dichloromethane (250 mL). Organic fractions were dried \textit{in vacuo} and subjected to solid phase extraction using Supelco-Discovery C\textsubscript{18} cartridges (5 g),
eluting with a step gradient of 40 mL MeOH/H₂O solvent mixtures (10%, 20%, 40%, 60%, 80%, 100% MeOH) and finally with EtOAc to afford seven fractions. The resulting fractions were dried in vacuo, resolubilized in 500 µL of dimethyl sulfoxide (DMSO), and transferred to deep-well 96-well plates for screening.

2.4.7. Fermentation and Isolation.

The producing organism, MMA 6B HVS/10A, was isolated from the lower intestine of a harbor porpoise, specimen LMLPP2011SEP29, under permit number 151408SWR2011PR00001:SMW. The strain was originally isolated on HVS medium (18 g agar, 50 mg nalidixic acid, 50 mg cycloheximide, 10 g starch, 1.7 g KCl, 0.01 g FeSO₄·7H₂O, 0.5 g MgSO₄·7H₂O, 0.5 g Na₂PO₄, 3.0 g KNO₃, 0.5 mg thiamine (vitamin B1), 0.5 mg riboflavin (vitamin B2), 0.5 mg nicotinic acid (vitamin B3), 0.5 mg pyridoxine HCl (vitamin B6), 0.5 mg p-aminobenzoic acid, 0.5 mg myoinositol, 0.25 mg biotin, 20 mg CaCO₃, 750 mL 0.2-µm filtered seawater, 250 mL Milli-Q H₂O). All vitamins were dissolved in stock solutions, filtered, and added after autoclaving. Frozen stocks of environmental isolates were grown on fresh Marine Broth plates (37.4 g Difco Marine Broth, 18 g agar, 1 L Milli-Q water) and incubated at 25 °C until discrete colonies became visible. Selected colonies were inoculated into 10 mL modified saline SYP (mSYP) media (10 g starch, 4 g peptone, 2 g yeast extract, 31.2 g Instant Ocean in 1 L distilled H₂O). The cultures were stepped up in stages at 7-day intervals by first inoculating 1.5 mL of the 10-mL cell cultures into 50 mL mSYP (medium scale), followed by inoculation of 40 mL of these medium-scale cell cultures into 1 L of the same broth also containing 20.0 g Amberlite XAD-16
adsorbent resin in 2.8 L Fernbach flasks for 7 days. All cultures were incubated at 25 °C containing glass beads for 10 mL cultures and stainless steel springs for 50 mL and 1 L cultures, and shaken at 200 rpm.

The cells and resin were collected from the bacterial extract by vacuum filtration using Whatman glass microfiber filters and washed with deionized water. This cell/resin slurry was extracted with 250 mL of 1:1 MeOH/CH$_2$Cl$_2$, and the organic extract was removed by vacuum filtration and concentrated to dryness in vacuo. The crude organic extract (code RLFI1036) was subjected to solid-phase extraction using a Supelco-Discovery C18 cartridge (10 g) and eluted using a step gradient of 80 mL MeOH/H$_2$O solvent mixtures (10% MeOH (A), 20% MeOH (B), 40% MeOH (C), 60% MeOH (D), 80% MeOH (E), 100% MeOH (F), and finally ethyl acetate (G)) to afford seven fractions designated as A–G. The seven fractions were dried in vacuo. The 100% MeOH prefraction, RLUS1036F, was subjected to reverse phase HPLC (Phenomenex Synergi Fusion-RP 5 µm, 80 Å, 250×4.6 mm, 65:35 MeOH/H$_2$O isocratic run over 20 min, 1 mL min$^{-1}$ flow rate) to produce phocoenaminicin, which eluted at 15.7 min.

2.4.8. Flow Cytometry.

Membrane potential was analyzed on a Becton Dickinson LSRII flow cytometer using the reagents and protocol from the BacLight™ Bacterial Membrane Potential Kit (B34950).
2.4.9. Bacterial Strains.


2.4.10. High-Throughput Antibacterial Inhibition Assay.

Bacterial test strains were grown on fresh agar plates and individual colonies were used to inoculate 3 mL sterile media. All *Staphylococcus* strains were grown in tryptic soy agar (TSA; 17 g tryptone, 3 g soytone, 2.5 g dextrose, 5 g NaCl, and 2.5 g dipotassium phosphate in 1 L distilled water; pH 7.5). *P. alcalifaciens, O. anthropi, E. aerogenes,* and *A. baumanii* were grown in nutrient broth agar (Difco, USA), and *B. subtilis, E. coli, V. cholerae, S. Typhimurium, P. aeruginosa,* and *Y. pseudotuberculosis* cultures were grown in Luria Broth (10 g tryptone, 5 g yeast extract and 10 g NaCl in 1 L distilled water; pH 7.5). *C. difficile* was grown on TSA plates supplemented with 5% sheep blood (Hardy Diagnostics, VWR catalog number 89405-024). All *C. difficile* media was placed in anaerobic chambers 24 h prior to
inoculation, and all culture conditions were made anaerobic using disposable chambers (BD Diagnostics, 90003-642). All in-house media were autoclaved at 121 °C for 30 min. Antimicrobial MICs were determined using Mueller Hinton Broth (VWR catalog number 90003-966). Inoculated cultures were grown overnight with shaking (200 rpm, 30 °C). Saturated overnight cultures were diluted 1:10000, 1:1000, or 1:100 according to turbidity and dispensed into sterile clear polypropylene 384-well plates (30 µL screening volume). Optical density (OD$_{600}$) of cultures at a 1:100 dilution were recorded (Shimadzu UV-Visible Spectrophotometer) and further diluted on agar plates to calculate colony forming units (CFU) per milliliter of culture. DMSO solutions of test compounds (200 nL) were pinned into each well prior to inoculation using a high-throughput pinning robot (Perkin Elmer Janus MDT). In the 384-well plates, lanes 1 and 2 were reserved for DMSO vehicle negative controls, while lanes 23 and 24 contained only culture medium and test organisms. After compound addition, screening plates were stacked in an automated plate reader/shaker (Perkin Elmer EnVision) and OD$_{600}$ was measured every 1 h for ~18 h. The resulting growth curves for each dilution series were used to determine MIC values for all test compounds following standard procedures.
Table 4 MIC values of phocoenamicin in variable pH and calcium concentrations.

<table>
<thead>
<tr>
<th>Ca Concentration</th>
<th>Phocoenamicin MIC (µM)</th>
<th>Vancomycin MIC (µM)</th>
<th>Daptomycin Mic (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ 0 mM Ca</td>
<td>3.5</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>+ 0.5 mM Ca</td>
<td>4.2</td>
<td>-</td>
<td>0.5</td>
</tr>
<tr>
<td>+ 1.25 mM Ca</td>
<td>4.2</td>
<td>-</td>
<td>0.1</td>
</tr>
<tr>
<td>pH 7.0</td>
<td>1.3</td>
<td>1.3</td>
<td>-</td>
</tr>
<tr>
<td>pH 6.0</td>
<td>0.63</td>
<td>0.78</td>
<td>-</td>
</tr>
<tr>
<td>pH 5.0</td>
<td>0.53</td>
<td>1.1</td>
<td>-</td>
</tr>
</tbody>
</table>

2.4.11. Cytological Profile Screening and Image Analysis.

Methods for cell culture and staining were used as previously reported (32, 33). HeLa cells were plated into two clear-bottom 384-well plates at a target density of 2,500 cells per well. The plates were incubated for 24 h under 5% CO\textsubscript{2} at 37 °C, 150 nL of extract was pinned onto the culture plates, and the plates were incubated for 19 h under 5% CO\textsubscript{2} at 37 °C. The plates were then fixed and stained with sets of either cell-cycle or cytoskeletal stains, which report on the number of cells in S-phase or mitosis and the amount and distribution of tubulin and actin, respectively. Both stain sets also contained a nuclear stain (Hoechst), which was used to count the number of cells and segment the image. The plates were imaged with a 10X objective lens, acquiring four images per well. For each extract, 248 different parameters were measured from the images of each plate. Together, these values report on a diverse range of size and shape features, such as the total area and shape of the nuclei and the number of mitotic cells. Comparing extract-treated and DMSO-treated wells and reduction of these cellular metrics to population values for each well using a custom data management pipeline produced a 248-parameter fingerprint for each extract with attributes scaled from -1 to 1 (Figure 2.9B).
Figure 2.9 Cytological profiling data. (A) Images of drug treated HeLa cells stained with Hoechst dye (DNA), anti-phosphohistone H3 antibody (mitotic marker), and EDU (clickable version of BrdU, a metabolically incorporated nucleoside analog used as an S-phase marker). Cells undergoing DNA synthesis are shown in green, mitotic cells in pink, and cell nuclei in blue. (B) Color bar used to generate fingerprints
2.4.12. Extended Structure Elucidation


White amorphous powder, $[\alpha]^{25}_D = -6.3$, (c = 1.0, Acetone); UV(MeOH) $\lambda_{\text{max}}$ (log $\varepsilon$) 232(3.34), 398(2.89), 314(2.82) nm; $^1$H and $^{13}$C see NMR table; HRMS [M-H]$^-1$ 1069.4571 (calculated for C$_{56}$H$_{74}$ClO$_{18}$, 1069.4564)

![Phocoenamicin, [M-H]$^-1$ 1069.4](image)

Figure 2.10 LCMS trace of crude extract, RLFH1036F. Spectra obtained using a gradient of MeCN:H$_2$O + 0.02% formic acid (65% MeCN for 2 min, 65%–95% MeCN over 12 min) at a flow rate of 2 ml/min (Phenomenex Synergi Fusion-RP, 10 x 250 mm column).

2.4.12.2. Subunit A (Glycosylation).

Initial evaluation of the $^1$H and $^{13}$C spectra, coupled with HRMS data, suggested a molecular formula of C$_{56}$H$_{75}$ClO$_{18}$. Examination of the HSQC spectrum suggested the presence of an aliphatic core, two sugar moieties, and a poly-substituted aromatic substituent (S3A). Evaluation of the COSY spectrum starting from the anomeric proton at 4.41 revealed a 6-deoxy-hexose sugar. HMBC from the 3-position (3.45 ppm) to a new anomeric carbon at 105.3 indicated a 1,3-linkage to a second sugar subunit. COSY correlations from this anomeric proton (4.61 ppm) revealed the presence of a second 6-deoxy-hexose sugar unit. HMBC from the proton at the 4
position (4.93 ppm) to a new quaternary carbon at 168.0 suggested the presence of a terminal substituent attached 1-4 to this sugar unit. HMBC correlations from the proton (4.93 ppm) on the 4-position of the second sugar to the carbonyl at 168.0 ppm indicated an acetyl linkage to the aromatic core. HMBC correlations between aromatic proton (6.85 ppm) to the same carbon also confirmed the linkage. COSY correlations suggested adjacent aromatic protons. Comparison of $^{13}$C chemical shifts to the known aromatic subunit of chlorothricin (S3C) as well as synthetic analogues of the aromatic core confirmed the regiochemistry of the ring. Together, these data identified subunit A.

2.4.12.3. **Subunit B (Trans decalin).**

To access the aliphatic core we started with the methylene protons at 1.91 and 2.30 ppm on C17. COSY correlations revealed a long linear spin system comprising ten contiguous protonated carbons (Subunit B1, S3A). HSQC and HMBC correlations confirmed these assignments, and revealed the presence of an additional subunit comprising seven carbons (Subunit B2). These two subunits were connected through key HMBC correlations between the protons on C25 to C13; and from the protons on C27 to the methylene at C7. Additional HMBC correlations identified a decalin core ring structure was identified via HMBC correlations from H5 (1.82) to C9 and C10 (87.6, 47.7 ppm)
2.4.12.4. Subunit C (Cyclohexene and diol sidechain).

Subunit C was identified starting with an olefinic proton at 5.05 ppm on C19, attached to a corresponding carbon at 131.0 ppm, C19 (S3A). HMBC to a second quaternary olefinic carbon at 134.3 suggested the presence of a tri-substituted olefin. HMBC correlations from a singlet methyl at 1.72/22.4 ppm to both olefinic carbons indicated a methyl substitution on C20, which was confirmed by a reciprocal HMBC correlation from the olefinic proton at 5.05 to the methyl carbon at 22.4 ppm. This subunit was extended through a combination of HMBC, COSY and TOCSY correlations to give a five carbon chain from C19 to C32 which was connected to a tertiary alcohol C33 via an HMBC correlation between the methyl protons on C36 and C32. This was further extended through HMBC correlations from the protons on the methyl C36 to a carbonyl carbon C34 at 212.8 ppm. This carbonyl belonged to a small terminal subunit comprised of a diagnostic methyl ketone single at 2.21/24.8 ppm. HMBC correlations between the methyl protons of C29 to the carbons at C19 and C23; and in between the methylene protons of C22 to C23 closed this subunit as a cyclohexene. Reciprocal HMBC correlations between C29 and C17 allowed the connection of subunit B and C between C17 and C18.
Figure 2.11. Structure elucidation of phocoenamicin. (A) Subunits for NMR analysis. (B) ROESY correlations. (C) The structure of chlorothricin has identical aromatic substitution as phocoenamicin.
2.4.12.5. Planar Structure Assembly.

To complete the planar structure assignment these subunits were connected as follows: Connect B and C: Complementary HMBC correlations from singlet methyl 29 to methylene carbon 17, and from methylene protons 17 to quaternary carbon 18 confirmed the presence of a C-C bond between C17 and C18, thus linking subunits B and C. The singlet methyl at 1.21 in subunit D showed a diagnostic HMBC correlation to the oxygenated methine carbon at 73.3 ppm in subunit C, thus linking the terminal subunit D moiety through a new C-C bond between C32 and C33. Finally, reciprocal HMBC correlations between the oxygenated methine at C9 on subunit B and the anomeric position at C1’ on the glycosidic sidechain revealed an ether linkage between these two motifs, completing the assembly of all the major fragments. With this extended fragment in place, consideration of the molecular formula indicated the absence of atoms C3HO2 and four DBEs. Comparison of this late-stage partial structure with the existing literature revealed a number of other molecules with similar core motifs. Common to these structures is a central tetronic acid motif that was internally consistent with the remaining atoms for phocoenaminicin. Installation of this functional group completed the planar structural assignment, which was further confirmed by consideration of the $^{13}$C chemical shifts for this final structural element.
2.4.13. Configurational Analysis

2.4.13.1. Aliphatic Core.

First, arrangement of decalin core was determined by examining ROE and selected $^3J_{HH}$ coupling constants that defined the relative configurations of all chiral centers in this 6-6 fused ring system (S3B). A key ROE between H13 and H15 defined H15 on the bottom face of the molecule. A large $^3J_{HH}$ coupling constant between H15 and H16 assigned the olefin as trans. Examination of the ROE correlations for H15 to CH$_3$ 28 and H16 to CH$_3$ 29 defined these two methyl groups as down and up respectively. A coupling constant of 10.8 Hz between CH$_3$ 29 and H17$\alpha$ suggested that they were on opposite faces. ROE correlations between CH$_3$ 29 to the methylene proton at 2.30 ppm on C22(H22$\alpha$) in addition to a ROE between H21 and H22$\alpha$ assigned H21 to the top face. Side chain configuration is discussed within the main text. While heteronuclear coupling experiments were inconclusive, based on the biosynthesis of this class of molecules we assigned the tetronic acid stereogenic center so that it aligned with the other spirotetronates.
Figure 2.12 ROE Correlations of Possible Diastereomers. A combination of coupling constants and key ROE correlations indicated a configuration of 32S, 33R for the sidechain. To further justify this conclusion, all possible diastereomers of the side chain were modeled in Discovery Studio 4.0 (i, ii, iii). Except for the (32S, 33R) diastereomer shown in purple (iv) all other diastereomers were unable to reproduce all the observed ROE correlations. In any one confirmation, at least one observed ROE correlation was unobtainable in all the other models created. Examples of ROE correlations that were observed but not able to be replicated in the model without sacrificing other correlations are represented with yellow double headed arrows. Correlations in the models consistent with the observed data are displayed with white double headed arrows.

i) SS Stereoisomer: Observed NOE correlation not reproducible in model in yellow
ii) SS Stereoisomer: Observed NOE correlation not reproducible in model in yellow

ii) RS Stereoisomer: Observed NOE correlation not reproducible in model in yellow
2.4.13.2. Sugars.

Relative configuration determined by ROE and $^{3}J_{HH}$. Absolute, and confirmation of relative through degradation and comparison with standards (NMR for constitution and OR for absolute).

2.4.13.3. Synthesis of Triacetate 6-deoxy-D-Glucopyranose.

Phocoenamicin (10 mg, 0.009 mmol) was subjected to methanolysis by refluxing in 4 mL 1N HCl:MeOH for 2 hours. Upon cooling, the reaction mixture was concentrated to dryness in vacuo and partitioned between H$_2$O (10 mL) and Et$_2$O (10 mL). The phases were separated and the aqueous phase extracted with Et$_2$O (2 x 10
mL). The aqueous portion was lyophilized to dryness. To a mixture of the crude lyophilized product (1 mg) and Y(CF₃SO₃)₃ (0.1 mg in dry DCM (5 mL) was added acetic anhydride (190 mL) at rt. After being stirred for 2h the solution was quenched with saturated sodium bicarbonate solution, and the product was extracted with DCM three times. The combined organic extracts were washed with water, and brine. The triacetate 6-deoxy-D-glucopyranose was purified using a Waters mass-directed preparative system equipped with an XBridge BEH130 5 µM 19 x 150 C18 column eluting with H₂O/Acetonitrile. triacetate 6-deoxy-D-glucopyranose was also synthesized from 6-deoxyglucopyranose using identical procedures. Optical rotations of the anomers were consistent with synthetic compound.
### 2.4.14. NMR Chemical Shifts

<table>
<thead>
<tr>
<th>Position</th>
<th>$\delta_H$ (J in Hz)</th>
<th>$\delta_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>175.0, qC</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>106.9, qC</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>200.3, qC</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>50.1, qC</td>
</tr>
<tr>
<td>5</td>
<td>1.82, m</td>
<td>42.6, CH</td>
</tr>
<tr>
<td>6</td>
<td>1.56, m</td>
<td>38.7, CH</td>
</tr>
<tr>
<td>7α</td>
<td>1.75, m</td>
<td>45.0, CH$_2$</td>
</tr>
<tr>
<td>7β</td>
<td>1.20, m</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>1.62, m</td>
<td>40.3, CH</td>
</tr>
<tr>
<td>9</td>
<td>3.07, dd (10.2, 10.2)</td>
<td>87.6, CH</td>
</tr>
<tr>
<td>10</td>
<td>1.94, 1.94, m</td>
<td>47.8, CH</td>
</tr>
<tr>
<td>11</td>
<td>6.40, d (10.0)</td>
<td>126.5, CH</td>
</tr>
<tr>
<td>12</td>
<td>5.59, ddd (10.0, 6.0, 2.4)</td>
<td>125.5, CH</td>
</tr>
<tr>
<td>13</td>
<td>2.67, m</td>
<td>42.6, CH</td>
</tr>
<tr>
<td>14</td>
<td>2.03, m</td>
<td>39.4, CH</td>
</tr>
<tr>
<td>15</td>
<td>5.43, ddd (15.0, 9.0)</td>
<td>144.7, CH</td>
</tr>
<tr>
<td>16</td>
<td>5.12, ddd (15.9, 10.8, 2.4)</td>
<td>121.9, CH</td>
</tr>
<tr>
<td>17α</td>
<td>2.30, dd (12.6, 10.8)</td>
<td>43.2, CH$_2$</td>
</tr>
<tr>
<td>17β</td>
<td>1.91, dd (12.6, 2.4)</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>40.6, qC</td>
</tr>
<tr>
<td>19</td>
<td>5.05, s</td>
<td>131.0, CH</td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>134.3, qC</td>
</tr>
<tr>
<td>21</td>
<td>2.4, br ddd (9.6, 7.4, 3.1)</td>
<td>33.7, CH</td>
</tr>
<tr>
<td>22α</td>
<td>2.30, dd (13.9, 7.4)</td>
<td>29.9, CH$_2$</td>
</tr>
<tr>
<td>22β</td>
<td>1.85, m</td>
<td>-</td>
</tr>
<tr>
<td>23</td>
<td>-</td>
<td>86.5, qC</td>
</tr>
<tr>
<td>24</td>
<td>-</td>
<td>204.9, qC</td>
</tr>
<tr>
<td>25</td>
<td>1.65, s</td>
<td>16.8, CH$_2$</td>
</tr>
<tr>
<td>26</td>
<td>0.83, d (6.6)</td>
<td>23.3, CH$_1$</td>
</tr>
<tr>
<td>27</td>
<td>1.05, d (6.6)</td>
<td>19.8, CH$_1$</td>
</tr>
<tr>
<td>28</td>
<td>0.86, d (7.2)</td>
<td>21.3, CH$_1$</td>
</tr>
<tr>
<td>29</td>
<td>1.27, s</td>
<td>24.4, CH$_1$</td>
</tr>
<tr>
<td>30</td>
<td>1.72, s</td>
<td>22.2, CH$_1$</td>
</tr>
<tr>
<td>31α</td>
<td>1.98, ddd (13.2, 9.6, 4.8)</td>
<td>33.6, CH$_2$</td>
</tr>
<tr>
<td>31β</td>
<td>1.75, ddd (13.2, 10.5, 3.1)</td>
<td>-</td>
</tr>
<tr>
<td>32</td>
<td>3.87, d (10.2, 4.8)</td>
<td>73.3, CH</td>
</tr>
<tr>
<td>33</td>
<td>-</td>
<td>82.5, qC</td>
</tr>
<tr>
<td>34</td>
<td>-</td>
<td>212.8, qC</td>
</tr>
<tr>
<td>35</td>
<td>2.21, s</td>
<td>24.9, CH$_1$</td>
</tr>
<tr>
<td>36</td>
<td>1.21, s</td>
<td>22.0, CH$_1$</td>
</tr>
<tr>
<td>1'</td>
<td>4.41, ddd (6.2, 2.4)</td>
<td>103.6, CH</td>
</tr>
<tr>
<td>2'</td>
<td>3.45, m</td>
<td>74.6, CH</td>
</tr>
<tr>
<td>3'</td>
<td>3.45, m</td>
<td>89.5, CH</td>
</tr>
<tr>
<td>4'</td>
<td>3.09, m</td>
<td>74.9, CH</td>
</tr>
<tr>
<td>5'</td>
<td>3.27, m</td>
<td>72.2, CH</td>
</tr>
<tr>
<td>6'</td>
<td>1.24, d (6.6)</td>
<td>18.2, CH$_1$</td>
</tr>
<tr>
<td>1&quot;</td>
<td>4.62, d (7.9)</td>
<td>105.3, CH</td>
</tr>
<tr>
<td>2&quot;</td>
<td>3.5, ddd (8.4, 8.4)</td>
<td>75.3, CH</td>
</tr>
<tr>
<td>3&quot;</td>
<td>3.78, t (9.4, 8.6)</td>
<td>74.7, CH</td>
</tr>
<tr>
<td>4&quot;</td>
<td>4.93, t (9.6, 9.6)</td>
<td>77.4, CH</td>
</tr>
<tr>
<td>5&quot;</td>
<td>3.82, dq (9.6, 6.0)</td>
<td>70.9, CH</td>
</tr>
<tr>
<td>6&quot;</td>
<td>1.33, d (6.6)</td>
<td>17.9, CH$_1$</td>
</tr>
<tr>
<td>7&quot;</td>
<td>-</td>
<td>160.0, qC</td>
</tr>
<tr>
<td>8&quot;</td>
<td>-</td>
<td>122.6, qC</td>
</tr>
<tr>
<td>9&quot;</td>
<td>-</td>
<td>155.6, qC</td>
</tr>
<tr>
<td>11&quot;</td>
<td>7.35, d (8.4)</td>
<td>132.6, CH</td>
</tr>
<tr>
<td>10&quot;</td>
<td>6.85, d (8.4)</td>
<td>116.2, CH</td>
</tr>
<tr>
<td>12&quot;</td>
<td>-</td>
<td>125.7, qC</td>
</tr>
<tr>
<td>13&quot;</td>
<td>-</td>
<td>135.7, qC</td>
</tr>
<tr>
<td>14&quot;</td>
<td>2.41, s</td>
<td>18.0, CH$_1$</td>
</tr>
<tr>
<td>OH</td>
<td>4.10</td>
<td>-</td>
</tr>
<tr>
<td>OH</td>
<td>3.92</td>
<td>-</td>
</tr>
<tr>
<td>OH</td>
<td>3.65</td>
<td>-</td>
</tr>
</tbody>
</table>
δ(pgm) gCOSY of phocoenaminic at 600 MHz in acetone-d6
$\delta_{(m,p)}$

gTOCSY of phocoenamicin at 600 MHz in acetone-d$_6$. 

$\delta$ (ppm)
HSQC of phocoenamicin at 600 MHz in acetone-d6
HMBC of phocoenaminic in acetone-d6.
ROESY of phocoenamin at 600 MHz in acetone-d6
$\delta_{ppm}$ of phocoenamicin at 600MHz in acetone-d$_6$
1D-TOCSY of phocoenaminic at 600 MHz in acetone-
6 irradiated at 2.45 ppm
1D-TOCSY of phocoenamicin at 600 MHz in acetone-d$_6$ irradiated at 5.1 ppm
2.4.15.1. $^1$H of synthetic and isolated sugars at 600 MHz in chloroform-d.
2.4.15.2. Synthetic and isolated sugar co-injection.

Isolated triacetate 6-deoxy-β-D-glucopyranose

Synthetic triacetate 6-deoxy-β-D-glucopyranose

Co-injection
2.5. References


(34) Avalos-Téllez, R.; Suárez-Güemes, F.; Carrillo-Casas, E. M.; Hernandez-Castro, R. Bacteria and Yeast Normal Microbiota From Respiratory Tract and Genital Area of Bottlenose Dolphins (*Tursiops Truncatus*). In *Current Research Topics in Applied Microbiology and Microbial Biotechnology*;


(40) Wang, H.; Claveau, D.; Vaillancourt, J. P.; Roemer, T.; Meredith, T. C.


(48) Felsenstein, J.; Fischbach, M. A.; Segre, J. A. Confidence Limits on


Chapter 3

Mining Genetic Interactions in *Saccharomyces cerevisiae* for Drug Discovery
# Table of Contents

3. **Mining Genetic Interactions in *Saccharomyces cerevisiae* for Drug Discovery** ................................................................. 109

3.1. **Introduction** ......................................................................................................................................................... 109

3.2. **Results/Discussion** ............................................................................................................................................... 116

3.2.1. Prioritization by betweenness centrality .......................................................................................................................... 116

3.2.2. Validation using Boone data ............................................................................................................................................. 119

3.2.3. Binning of genes using Gene Ontology ........................................................................................................................... 122

3.2.4. "Hub gene" selection and ranking ................................................................................................................................. 126

3.2.5. Comparison of acquired data to published data ............................................................................................................. 131

3.3. **Conclusion** ......................................................................................................................................................... 133

3.4. **Experimental** ................................................................................................................................................... 134

3.4.1. Cell culturing ......................................................................................................................................................... 134

3.4.2. Data Analysis ......................................................................................................................................................... 134

3.5 **References** ................................................................................................................................................... 137
3. Mining Genetic Interactions in Saccharomyces cerevisiae for Drug Discovery.

3.1. Introduction

Mechanism of action (MOA) annotation is still one of the most challenging and time consuming aspects of drug discovery. Cytological profiling is a high content assay that can rapidly generate large amounts of data for libraries of pure compounds or complex mixtures. The strength of this technology lies within the capability to identify the MOA of unknown compounds by using a pattern matching algorithm to align them with compound treated wells possessing similar MOAs. While this has proven successful in the past, it has an inherent limitation. “Unknown clusters” found in the library that do not align with any known training set compounds are placed randomly and exhibit limited overlap when compared to their neighboring clusters. These may be unique compounds with unusual modes of action. However, it is also likely that the set of compounds possess a known MOA not included in the library training set. Compounds possessing unique MOAs are unable to be characterized in the cytological profiling assay, and need an orthogonal method not limited by the training sets used to validate the assay. Generation of a training set-independent assay would circumvent this problem, segregate the unknowns and help mitigate the rediscovery of known chemistry.

Proteomics approaches have been successfully applied to the determination of small molecule MOAs. While these methods have had great success in annotating the mechanism of small molecules, this process typically involves chemical
modification of the small molecule of interest to detect the protein target. Chemical manipulation of compounds can modulate or abolish the bioactivity altogether and may not be a true representation of the molecule in its native form. Data interpretation of these experiments is not trivial and often times abundant proteins with low affinity for the compound will generate significant noise during the analysis. A quantitative proteomics approach using isobaric labeling (iTRAQ) or single isotope labeling (SILAC) can label a drug treated proteome so that proteins specifically affected by the drug can be identified. This approach offers global proteome coverage without chemical modification to the small molecule. However, without an enrichment step, such as affinity purification used with labeled small molecules, resolving the entire proteome with the added complexity of isotopic or isobaric labeling via LCMS/MS remains a challenging aspect of proteomics. Each run can take several days to optimize, to prepare fractionated samples, and to run. In addition, considering that natural product extracts are of unknown composition and low titer, traditional pull-down experiments that require synthetic manipulation of the small molecules are not generally feasible and this proteomic analysis must therefore be accomplished with a label free proteomic methodology.

Previous studies have shown that the binding of small molecules to proteins can stabilize folded conformations and provide protection from degradation upon proteolysis. Drug affinity responsive target stability (DARTS) is a technique applied by the Huang group to identify discrete protein targets. The proteolysis step enriches the analysis for proteins protected from degradation through small molecule binding
and therefore simplifies the analysis. This method has yet to be used on a large scale, and is also limited by any protein targets intrinsically sensitive to proteolysis. Overall, proteomics is a powerful tool for annotating protein targets one compound at a time; however, it is not currently amenable to large sets of compounds and complex mixtures.

While not all proteins have been associated with gene targets, genomic studies offer an alternative method for the analysis of compound MOAs on a larger scale without synthetic manipulation. A forward chemical genetics approach identifies compounds that elicit a desired phenotype, followed by identification of the target through affinity pull-down assays or sequencing of resistant mutants.\textsuperscript{10} If the compound has a discrete target, then comparing the sequencing data from the resistant and original organism can lead to the identification of specific gene targets.\textsuperscript{11} One challenge with this method is that many organisms can mutate to activate their efflux pumps to expel the toxic drug; these types of mutations are not useful for identifying the drug mechanism as seen with phocoenamicin in chapter 2. Reverse genetic approaches use gene deletions or other genetic perturbations to study a phenotype derived from a specific mutation.\textsuperscript{12} This has been performed on entire genomes of organisms such as \textit{Saccharomyces cerevisiae} (yeast), \textit{Saccharomyces pombe}, and \textit{Caenorhabditis elegans}, and is an expanding field of research.\textsuperscript{13-15}

Some genetic based systems have successfully incorporated mammalian cells in a method independent of a compound training set use RNAi and CRISPRi techniques to generate desired genetic manipulations.\textsuperscript{16,17} The Functional Signal
Ontology (FUSION) system developed by researchers at the University of Texas Southwestern Medical Center shows that the careful selection of highly variable and diagnostic gene expression signatures can annotate the modes of action of unknown molecules in a manner that can identify targets that are not currently associated with small molecules. This technology has demonstrated DDR2 as the molecular target of a new family of alkaloid natural products, discoipyrroles A–D. The only weakness is that this method isn’t available for widespread lab use, as it is expensive to acquire and maintain all genetic perturbations, and it requires the use of QuantiGene Plex 2.0 assays (Panomics) with a Luminex 200 machine (Luminex) at an approximate cost of greater than $60,000. Therefore, rapid and cost effective MOA annotation of small molecules remains a challenging aspect of drug discovery.

Although it would be ideal to study small molecules in mammalian cells, they are significantly complex, making MOA annotation in a simpler system appealing. The yeast proteome consists of approximately 6,000 possible protein targets compared to mammalian cells which possess over 30,000 different protein targets, many of which have not been characterized. As one of the most studied organisms, *S. cerevisiae* has a growing knowledgebase with regards to the entire genome and proteome. A consortium of geneticists has formed a large deposit of public genetic information on the Saccharomyces Genome Database compiling genetic interactions acquired from multiple studies, making it readily available to researchers. *S. cerevisiae* has been heavily studied and annotated and has proven to be a model organism for drug discovery as it contains a remarkable amount of conserved
properties with higher organisms. The ease of manipulation and handling makes *S. cerevisiae* an ideal candidate for studying the bioactivity of drug candidates.

Perturbations of every gene in the *S. cerevisiae* genome have been extensively studied by many labs including the Boone and Giaever labs, revealing information about the function and connectivity of the genome. These studies show that only ~20% of yeast genes are essential for survival. Essential genes can be analyzed through the use of conditional alleles. An exhaustive study on the overall fitness of double deletion mutants created a network of genetic interactions for the entire genome. Specifically, synthetically lethal interactions have been carefully observed. Synthetic lethality is observed when two specific deletions on their own do not cause cell death, but the combination of the two gene deletions is fatal.

Many reverse genetic techniques have been extended with the use of small molecules. While two genes can have a synthetically lethal interaction, this interaction can also be mimicked with small molecules. For a pair of synthetically lethal genes referenced as gene X and gene Y, if one of the gene deletions is replaced by a small molecule that targets gene X, when it is paired with a gene Y deletion mutant, that interaction will also compromise cell viability, Figure 3.1. Capitalizing on this principle, if a small molecule is screened against a library of single gene deletion mutants, then perturbations in fitness compared to the compound treated wildtype organism can reveal the compound’s MOA. Overexpression libraries can identify the gene target when overexpression of the gene target leads to rescue of the
organism. With synthetic lethality, a network is generated using the fitness of drug treated mutants to identify potential gene targets. All possible combinations of double gene deletion mutants were created, and the fitness of those mutants was analyzed for positive and negative interactions. These data create a network, which maps the genetic interactions and shows that genes of similar biological processes tend to possess similar interaction profiles. The Boone lab has evaluated many global and partial networks generated by synthetic genetic array analysis. Analysis of discrete biological classes as determined by the saccharomyces genome database (SGD) revealed that they were enriched with several negative genetic interactions. Processes related to chromatin and transcription were some of the most highly connected clusters in the network, and bridged several other biological processes, compared with protein interactions which share limited interactions between biological processes.

Figure 3.1 Synthetic Lethality. On left synthetic lethality is demonstrating using double gene deletions. On right, synthetic lethality is demonstrating using a single gene deletion and a drug that mimics the gene X deletion.
Large institutes, such as the Broad Institute of Harvard and MIT, screen small molecule libraries against the entire genomes of single deletion strains. This has proven useful in determining gene and protein targets of small molecules, and has added insight into off target effects. For example, this approach has demonstrated that natural product papuamide B targets phosphatidylserine in yeast, and that the marine sponge metabolite girolline targets Elongation Factor 2, and therefore exerts its anti-inflammatory activity through inhibition of protein synthesis at the elongation step. However, this process requires over 6,000 yeast strains to be cultured and analyzed for each compound of interest. Including the essential genes takes extra care to maintain and screen since they are not viable as single gene deletions. Small molecule targets can be identified in separate ways using synthetic genetic arrays. The first method developed uses monocultures where each gene deletion is cultured individually and queried against 12 training set compounds on agar plates. With the advancement of sequencing technology, DNA barcodes allow mass culture of pools of mutants to simplify this procedure. This process is supposed to correct for deviations between different strains with varying degrees of fitness. However, comparing pooled data and individual growth data shows some deviation between the two methods. While effective, culturing and analyzing 6,000 genetic mutants is a tool limited to geneticists and a small subset of compounds.

This chapter will address one methodology aimed at addressing this bottleneck of MOA annotation. To expedite the discovery of compounds with novel targets it was hypothesized that through a combined approach of careful gene
selection, similar to the Fusion platform, and using the extensive genetic network of *S. cerevisiae*, a small subset of gene deletions could be prioritized to determine broad MOA information for compounds in a cost-effective manner. This methodology consisted of iteratively prioritizing query genes through a variety of metrics such as betweenness centrality, interaction count and biological class specificity. Through each evaluation, prioritized query genes were selected from published data sets to generate synthetic genetic arrays with the aim of identifying the degree of localization for biological process to discrete gene sets. This analysis demonstrated that a set of query or hub genes dramatically smaller than the genome itself (<4 %) could be used to achieve modest genome coverage, and retain comparable resolution for many of the control compounds using published data-sets. Once optimized *in vitro*, this methodology could be utilized at any university interested in analyzing the biological activity of small molecules.

3.2. Results/Discussion

3.2.1. Prioritization by betweenness centrality

Through analyzing publically available data on the yeast genome from the SGD website, genes with the highest number of genetic interactions have been prioritized to maximize coverage of the genome. From the 6,000 genes, approximately 1,000 are essential or multidrug resistant and were therefore removed from our analysis. Removing the 2327 genes that have fewer than 5 genetic interactions gave a pool of 1122 remaining genes. The first metric used for prioritization was betweenness
Centrality is described by Freeman as “when a particular person in a group is strategically located on the shortest communication path connecting pairs of others, that person is in the central position.” For example, point P3 would be considered a central point because it forms the shortest connection between P1, P2, and P4 to P5 (Figure 3.2). Betweenness is defined as the ability of a node (ex. P2) between two other nodes (ex. P1 & P3) to facilitate or block communication: while P1 and P3 share partial control over the connection between P2 and P4, P3 holds complete access to P5. Through ranking those types of relationships, a network of the current data was assembled (Figure 3.3).

Through selection of 50 genes with the highest values for betweenness centrality, those genes had interactions with 62% of the non-essential genome. This coverage is defined as the percentage of genes...
that have at least one synthetic lethal interaction with the chosen subset of 50 genes. The number 50 was selected because it was the highest gene count the screening system workflow could handle in approximately one week. This low number was predicted to give modest coverage of the genome at a dramatically lower labor cost for cell culture and analysis when compared to the entire genome.

A vignette of the screening results is shown in Figure 3.4, in which 20 compounds with known mechanisms of action were screened against the 50 gene deletions strains

![Figure 3.4 Experimental screening results. X-axis: yeast deletion strains, Y-axis: dilutions of control compounds. Green = Negative growth, Black = no deviation in growth, Red = positive deviation in growth.](image-url)
with the highest betweenness score. The results were not as expected. Upon analysis, few gene deletions were sensitized to the drug, and many strains were unaffected. However, it appeared that many of the chosen strains were either resistant to the drugs altogether, or show MIC profiles similar to that of the wildtype strain, leading to inconclusive results. This preliminary round of screening showed that the initial plan for gene selection needed to be reconsidered. There is a fair amount of literature precedence documenting many yeast strains as multidrug resistant, and those would need to be precluded from our analysis. The selection of genes with the highest betweenness scores meant we were selecting genes with the greatest number of interactions in an attempt to get the largest genome coverage. However, while this might result in a higher hit rate in the assay, the results weren’t as meaningful. Getting a hit against a strain that has immediate interactions with every pathway in the genome doesn’t narrow down the MOA annotation, and does not drive the clustering based on biological processes. Therefore, this route of prioritization needed to be completely restructured to segregate hits into discrete bins of bioactivity. Prior to any additional screening, the genes selected should be optimized and visualized with published datasets to validate them as screening candidates.

3.2.2. Validation using Boone data

To determine the feasibility of this project, the genetic data acquired from 12 control compounds on agar plates for the entire genome by the Boone lab was used to validate the MOA analysis of small molecules with a reduced subset of genes. Initially the 50 genes previously picked for screening were investigated in silico
A limited number of interactions with the control compounds were observed (Figure 3.5A), and when the genes that clustered closest with the control

**Figure 3.5** Clustering of 50 genes with highest betweenness centrality scores. (A) Selected 50 genes (X-axis), 12 control compounds, entire genome (Y-axis). (B) Expansion of benomyl clustering and SGD enrichment results of closest genes in cluster. (C) Expansion of hydroxyurea and camptothecin cluster and SGD enrichment results of closest genes in cluster.
compounds such as benomyl (Figure 3.5B) and hydroxyurea (Figure 3.5C) were queried against the SGD database, no biological pathways were enriched, suggesting that the clustering was unable to annotate the MOA. Expansion of the analysis to the top 500 genes with the highest betweenness was performed *in vitro* (Figure 3.6). Evaluation of 500 genes demonstrated the feasibility of this project as many compounds clustered with their biological class (Figure 3.6A-B). Some were matched
with their gene targets, such as benomyl and CIN1 (Figure 3.6C),\textsuperscript{36} rapamycin and TOR1,\textsuperscript{11} and fluconazole and ERG11 (Figure 3.6D).\textsuperscript{30} Accurate MOA annotation using 500 genes versus 6,000 is a significant improvement, but still very time consuming to replicate in the lab without sophisticated robots and extensive culturing. These data show that it is possible to triage the entire genome to selective gene hubs, but additional analysis is required to determine if screening less than 500 genes can give necessary resolution.

3.2.3. Binning of genes using Gene Ontology

An additional resource of information on the SGD website is the annotation of the entire genome with Gene Ontology (GO) terms. GO terms have been used to annotate genes and gene products with standardized descriptions of their biological roles and functions.\textsuperscript{37} This annotation has been applied to over 347,778 species (including

![Gene ontology network for S. Cerevisiae](SGD 2017-02-25)

Figure 3.7 Gene ontology network for \textit{S. Cerevisiae}
strains) and is now an accepted method of gene annotation. Within the GO terms, there are many layers of annotations, and the GO terms themselves create a network as shown in Figure 3.7. This is only a simplified version of the GO term network, which is as complex as the genome itself; for example, the GO term for RNA accumulation has 495 GO terms associated within that node.

Many GO term categories could have been chosen for this analysis. The GO slim terms were selected because those categories were equally represented by both the essential genes and the non-essential genes, demonstrating the ability to test for most of the pathways represented by the genome. The GO slim category is a concise list of terms containing 101 GO slim terms which allows for broader binning (Figure 3.8). Upon close analysis of one of the larger GO slim terms, DNA repair, the relationship between selectivity within the compound class vs the count of genetic interactions

![Figure 3.8 Distribution of GO terms throughout the genome. X-axis: % of genes, Y-axis: GO terms, Red: whole genome, Blue: non-essential genome.](image)
showed that in general, the genes that have the highest selectivity (as defined by primarily having genetic interactions within a specified GO slim category compared to the rest of the genome) have the lowest amount of genetic interactions. With this information, a new hypothesis was formulated where genes could be selected to serve as markers for individual biological processes by looking at both the amount of genetic interactions, and how specific those interactions are to a biological process. An idealized version of the results is displayed in Figure 3.9, which show that a positive correlation between a compound with genes in “biological process A” would suggest that compound’s MOA is related to “biological process A.” With this hypothesis, the genome was analyzed for the existence of “marker genes” possessing a high enough selectivity and interaction count to be feasible for compound screening.
Figure 3.9 Idealized screening results. (i) Natural product isolation from microbes. (ii) Screening of compounds against a select set of genes for different biological processes. (iii) Compound 3 is associated with genes from biological process A for antimitotics, therefore compound 3 is likely an antimitotic compound.
3.2.4. “Hub gene” selection and ranking

For the top 50 biological processes containing the most genes, all genetic interactions within and outside their biological process were tabulated. Then data sheets were assembled so that genes could be selected through weighting the selectivity and interaction count for each biological process until an optimal result was formed. This was done by looking at ten biological processes at a time. In Figure 3.10, square A shows an NxN matrix of ten genes for ten biological processes. Currently shown are the top ten genes that have the highest interaction number per biological class. Note that each biological process has genes with multiple interactions throughout the

Figure 3.10 NxN matrix of the top ten GO terms. The GO term categories containing highest number of genes were chosen. Grey denotes no interaction. Red denotes lethal interactions. (A) Genes for each category were selected by highest betweenness centrality. (B) Genes for each category were optimized based on their interaction count and selectivity for having interactions primarily within their own class.
processes. Square B in Figure 3.10 shows how these interactions can be localized to different biological processes through weighting the selectivity of different genes. An NxN matrix is not shown for genes with the highest selectivity value because that consists of genes with one or few genetic interactions, and are statistically unlikely to interact with unknown compounds.

Once this was optimized for each biological process in batches of ten, those genes were combined and used to generate an NxN matrix shown in Figure 3.11 that represents the GO slim terms. While complete localization could not be achieved when compared to the entire genome, genes could be selected that had enriched interactions within their own biological class. The most time was put into optimizing

Figure 3.11NxN matrix of genes binned by GO terms.
genes from the DNA repair class, because it contains a large number of genes and is a challenging compound class to resolve with cytological profiling. These selected genes were then queried against the entire genome, and the interactions for some of the control compounds previously published by Boone were used to test the viability of this method of analysis (Figure 3.12). The top 17 genes from this category were carefully selected and when the entire genome was run against those selected genes, most of the genes did not have any interactions within these 17 (Figure 3.12A). Compounds that don’t interact with that process, such as fluconazole and tunicamycin, also did not have any interactions, which is to be expected (Figure 3.12D, F, H, I). The two DNA damaging compounds, camptothecin and hydroxyurea, both clustered closely together, and their ten closest genetic neighbors were run against the SGD library of GO terms (Figure 3.12G). The top hits were enriched for DNA replication, chromatin, or telomerase function and maintenance (Figure 3.13B).
Figure 3.12 Heatmap of top DNA damaging genes. (X-axis) and Boone control compounds clustered with the entire genome (Y-axis).

Wortmannin had one interaction with this subset of genes (Figure 3.12A). When all the surrounding neighbors were investigated, no GO terms were enriched, suggesting the BUB1 interaction was nonspecific for that cluster. Benomyl, a microtubule inhibitor isn’t a DNA damaging agent, yet it still clustered with its gene targets CIN1(Figure 3.12B). When the surrounding cluster was analyzed, an
enrichment for GO terms related to tubulin and mitosis were revealed (Figure 3.13). Considering that chromatin segregation and replication is an integral part of mitosis, it is not surprising that these biological processes are challenging to completely segregate. However it is very promising that the mitotic cluster was formed independently from the DNA damagers with so few genes tested.

Figure 3.13 Enriched GO terms. (A) Enrichment data for closest genes to benomyl. (B) Enrichment data for closes genes to hydroxyurea and camptothecin.
3.2.5. Comparison of acquired data to published data

While the clustering data generated from published data sets looked very promising, suggesting that smaller gene sets could be used to annotate MOA, these data were not consistent compared to the preliminary experimental results. Data curated in the lab were compared with the 12 compounds tested in individual strains; in general there were little to no similarities between the molecules tested (Figure 3.14). Camptothecin and hydroxyurea did not show the same sensitivity to the wildtype strain or mutant collection (from the Kellogg lab at UCSC) when compared to published results. There are various possibilities for this. Many compounds tested did not show the proper sensitivity to the wild type strain, suggesting that there could be some factor in the screening protocol affecting the outcome. This could include impure or degraded compounds in addition to inoculum effects changing the MIC values of the small molecules. To standardize this, colony forming unit counts should be acquired of all mutant strains to be tested and specific dilution procedures should be developed for each strain of interest. An additional possibility is that the integrity of our gene deletion library has been compromised.

The amenability of yeast to mutations makes it both excellent and challenging for genetic experiments. To test the integrity of our gene deletion library, the DNA bar codes for each strain should be checked to establish if there has been cross contamination, followed by complete sequencing to see if any unexpected mutations have arisen that could make the gene deletions less sensitive to drug treatment. There is also a chance that this difference can be attributed to processing the raw data
differently, or from using different mediums (agar to liquid). The BioMAP antibiotic profiling platform clusters and compares compounds of varying MIC values in different strains by normalizing them to the largest MIC value. However, normalizing with respect to the wild type activity or with respect to the largest MIC value did not greatly affect the data output.

**Figure 3.14** Comparison of experimental data (Black) with published data (Blue).
3.3. Conclusion

Analysis of the Boone data shows that a small subset can be used as an alternative to an entire genome based assay to give similar results. Analysis of 500 genes compared to the entire genome yielded remarkable resolution and clustered some compounds with their gene targets. GO slim terms are a useful method to categorize and annotate the yeast genome. Genes binned in different gene categories can be effectively prioritized as hub genes with an enriched specificity to their biological class. While a subset of less than 100 genes was not identified that could achieve global annotation of most biological processes, highlighting specific processes such as the DNA repair class can achieve selective MOA annotation. Moving forward, optimizing the processing of the data output, using inoculum counts verified by colony forming units, and acquiring certified strains or confirming strains by sequencing would serve to generate a cost effective and easily transferable screen that could efficiently bin small molecules into selected biological processes. Replication of the *in-silico* results in a screening center holds the potential of rapid MOA annotation in a time and cost effective manner.
3.4. *Experimental*

3.4.1. *Cell culturing*

Strains were grown on fresh agar plates and individual colonies used to inoculate 3 mL of sterile media (YPD, Fischer Scientific). Inoculated cultures were grown overnight with shaking (200 rpm; 30 °C). Saturated overnight cultures were diluted 1:1000. DMSO solutions of test compounds (200 nL) were pinned into each well at t₀ using a high-throughput pinning robot (Perkin Elmer Janus MDT). In the 384 well plate lanes 1 and 2 were reserved for DMSO vehicle negative controls, while lanes 23 and 24 contained only culture medium and test organisms. After compound addition, screening plates were stacked in an automated plate reader/shaker (Perkin Elmer EnVision) and a OD₆₀₀ reading was collected every 1 h for 16 - 20 h. The resulting growth curves for each dilution series were used to determine MIC values for all test compounds following standard procedures.

3.4.2. *Data Analysis*

Data from interacting gene pairs were assembled from published high-throughput yeast-two-hybrid results. The published set of genetic interactions was downloaded from BioGRID (www.thebiogrid.org, on May 2, 2015), and restricted to negative interactions and further limited to synthetic lethality. Genes were assigned one GO Slim term as determined from the GO term database on the SGD website (http://www.yeastgenome.org) on August 22, 2015. Betweeness centrality scores were calculated using Ulrik Brandes algorithm for betweenness centrality. Data taken
from the supplemental material for control compounds arrayed against ~4,700 haploid deletion single gene deletion strains individually cultured on solid media were incorporated into the genomic analysis used as controls for the *in silico* analysis.29

3.4.2.1. **Network Visualization**

Network figures were created with either the Cytoscape network visualization software using the perfuse force directed layout algorithm or using Gephi 0.8.2 and the force atlas2 layout algorithm.

3.4.2.2. **Two-dimensional Hierarchical Clustering**

The Hierarchical clustering was performed using Cluster 3.0 using an average linkage and Pearson correlation for distance metric. Clustering was visualized using Java Treeview and genes to compound clusters were selected by inspection and used to generate gene sets. These gene sets were submitted to the list analysis tool which includes gene ontology enrichment analysis from the SGD website (http://www.yeastgenome.org) on February 8th, 2017. This enrichment analysis makes use of Holm-Bonferroni method for which a max p-value of 0.05 was chosen as a threshold for enrichment of biological process ontology terms.

3.4.2.3. **Query Gene Selection**

Query genes were ultimately selected through evaluating each biological process separately and ranking the genes within that set based on the number of genetic interactions each gene had, and how selective those interactions were for within that biological process. Selectivity (S) was defined for each gene as the
percentage of the number of genetic interactions held with genes within its own biological process \((I)\) divided by the total number of interactions with the whole genome \((I_{tot})\).

\[ S = \frac{I}{I_{tot}} \times 100 \]

Each gene was ranked according to a generated metric \((M)\) determined by its weighted selectivity with regards to the average selectivity \((S_A)\) and interaction count \((I_A)\) for all genes within its specified biological process. Gene sets for each biological process were individually ranked through adjusting the weighting value \((W)\) between 0.5-5 to best prioritize the most genes possessing both a high selectivity within their biological process and a significant amount of interactions, based on the hypothesis that they were more likely to interact selectively when challenged with compound treatment.

\[ M = \left(\frac{I}{I_A}\right) + W \times \left(\frac{S}{S_A}\right) \]

In general, for each gene set, the weighting value \((W)\) was increased when comparison of an NxN matrix of many gene sets showed that the gene set of interest showed poor localization within their biological class. The weighting value \((W)\) was decreased when the heatmap revealed limited interactions for the genes within that gene set to favor genes with a higher interaction count over selectivity.
3.5. References


(14) Lopez, A.; Parsons, A. B.; Nislow, C.; Giaever, G.; Boone, C. Chemical-Genetic Approaches for Exploring the Mode of Action of Natural Products.


(17) Gonsalves, F. C.; Klein, K.; Carson, B. B. An RNAi-Based Chemical Genetic Screen Identifies Three Small-Molecule Inhibitors of the Wnt/Wingless Signaling Pathway; 2011.


(3), 319–327.


