VIDAR, WARREN S. Ph.D. Interaction Metabolomics to Discover Synergists in Natural Product Mixtures. (2022)
Directed by Dr. Nadja Cech. 64 pp.

Mass spectrometry metabolomics has become increasingly popular as an integral aspect of studies designed to identify active compounds from natural product mixtures. Classical metabolomics data analysis approaches do not consider the possibility that interactions (such as synergy) could occur between mixture components. With this study, we developed "interaction metabolomics" to overcome this limitation. The innovation of interaction metabolomics is the inclusion of compound interaction terms in the data matrix. The interaction terms are calculated as the product of the intensities of each pair of features (detected ions). Herein, we tested the utility of interaction metabolomics by spiking known concentrations of an antimicrobial compound (berberine) and a synergist (piperine) into a set of inactive matrices. We measured the antimicrobial activity for each of the resulting mixtures against *Staphylococcus aureus* and analyzed the mixtures with liquid chromatography coupled to high-resolution mass spectrometry (LC-MS). When the dataset was processed without compound interaction terms (classical metabolomics), statistical analysis yielded a pattern of false positives, a phenomenon that can be explained by confounding, in this case due to a left-out interaction in the model. However, interaction metabolomics correctly identified berberine and piperine as the compounds being responsible for synergistic activity. Our results demonstrate the utility of a conceptually new approach for identifying synergists in mixtures that may be useful for applications in natural products research and other areas that require comprehensive mixture analysis.

INTERACTION METABOLOMICS TO DISCOVER SYNERGISTS IN NATURAL PRODUCT MIXTURES

by

Warren S. Vidar

A Dissertation
Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro

2022

Approved by

Dr. Nadja B. Cech
Committee Chair

DEDICATION

This work is dedicated to my research mentor, to my family, and to my friends.

APPROVAL PAGE

This dissertation written by Warren S. Vidar has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair

Dr. Nadja B. Cech

Committee Members

Dr. Olav M. Kvalheim

Dr. Nicholas H. Oberlies

Dr. Daniel A. Todd

Dr. Qibin Zhang

September 21, 2022

Date of Acceptance by Committee

September 21, 2022

Date of Final Oral Examination

ACKNOWLEDGEMENTS

This work was supported by funding from that National Institutes Health National Center for Complementary and Integrative Medicine (NIH NCCIH) under grant number U41 AT008178 to RGL and NBC and grant number R15 AT010191 to NBC.

I would like to express my heartfelt gratitude to my mentor, *Dr. Nadja Cech* for putting so much confidence in me as her student and molding me to become a better scientist. I am most grateful for all the opportunities and guidance that she provided to me for the past four years. I thank her for trusting and believing in me especially during the tough times of research depression. She is such an amazing and inspirational mentor, and I will cherish her as a lifetime friend.

To the members of my committee, *Dr. Olav Kvalheim*, *Dr. Nick Oberlies*, *Dr. Daniel Todd*, and *Dr. Qibin Zhang*, thank you for lending your expertise and guidance to this research. Special thanks Dr. Zhang for being the first person to acknowledge my potential during the early phase of my application to UNC Greensboro.

To our collaborators, *Dr. Roger Linington* and *Dr. Tim Baumeister*, thank you for your significant contributions to this research.

To my friends, *Rodell Barrientos* and *Kristine Tolentino*, thank you for being my bridge to UNC Greensboro. Special thanks to Rodell for guiding me during the application process and for assisting me when I moved to the US.

To *Lindsay Caesar*, *Diane Wallace*, *Josh Kellogg*, *Heather Winter*, *Tyler Graf*, and *Herma Pierre*, thank you for dedicating your time to train and assist me in the lab.

To the *past and present members Cech Lab* and to my *cohorts in the department*, thank you for your support and for the fun times that we shared and spent together for the past years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

A central challenge in natural products research is to identify biologically active compounds in complex mixtures.[1–5] The gold standard approach towards accomplishing this task is bioassay-guided fractionation, wherein the mixture is subjected to successive stages of purification and biological evaluation until active compounds are identified. The value of bioassay-guided fractionation is evidenced by its history of success; many of the most therapeutically important natural products, molecules like artemisinin, Taxol, and penicillin, were discovered using this approach.[6,7] What happens, however, when the activity of a mixture is not due to a single compound, but to a mixture of compounds, which could act together synergistically, additively, or antagonistically? This question often arises in the study of botanical (herbal) medicines, which are employed therapeutically as mixtures rather than single molecules. Many proponents of the use of botanical medicines argue that they are effective by virtue of the combined action of multiple compounds.[8] A number of studies do point to the occurrence of synergistic biological effects in botanical extracts.[8–10] In a few cases, the specific constituents or mechanisms responsible for this synergy have been identified. For example, artemisinin has been shown to be more potent in vivo against malaria when used as a complex tea than as an isolated molecule,[11] and some plants contain both the antimicrobial alkaloid berberine **(Figure R1)** and additional molecules that enhance the activity of berberine via efflux inhibition.[5] However, the vast majority of natural product research focuses on the isolation of single active compounds, and there is a dearth of literature citing specific constituents that interact synergistically. It is possible that scenarios where multiple constituents in natural product mixtures exert meaningful combined biological activity are not, after all, very common. Alternately, perhaps our lack of knowledge about how combination effects arise is due to

limitations in our ability to study them. Approaches that focus on isolation and purification of single compounds may not fully explore the potential interactions that could contribute to the activity of mixtures.

There are five major requirements for an experimental design that enables identification of synergists in a mixture based on their association with biological activity. (1) Multiple mixtures must be evaluated for biological activity. (2) The active components must vary in concentration across the mixtures; otherwise, no new information is gained by testing multiple mixtures. (3) Two compounds that interact synergistically must be present at the correct range of concentrations to observe a synergistic effect. (4) The biological assay used must be appropriate for detection of synergy. (5) The method used to measure the presence and abundance of the mixture components must be able to detect the active constituents. Given requirements 1-5, there are multiple scenarios in which an analyst performing natural products drug discovery might fail to detect the presence of a synergist. The presence of a synergist will be missed if there are not enough measurements of biological activity, if the wrong biological activity is being measured, if the synergist and the active compound are not present in the same samples, if the synergist and the active concentration are not present at the correct concentrations to observe synergy, or if the analytical technique used for detection misses either the synergist or the active compound. Because of these inherent limitations, it will not be possible in a typical natural products drug discovery experiment to answer the question, “Is a synergist present in this natural product extract)?” The question that can be answered is the following question, “Could the activity that has been observed for a series of natural product mixtures be due to synergy between detectable compounds?” Herein, we sought to develop a metabolomics data analysis approach that would address this second question.

The most widely used and validated approach for studying interactions between two biologically active molecules is the checkerboard assay.[1,3,10] To conduct this assay, two compounds are tested in combination over a range of concentrations by two-fold dilution. The data are then plotted in the form of an isobologram, which visually represents the changes in dose-response behavior resulting from the combined effects of the two samples. If the dose-response behavior doesn't change when the two compounds are combined, the compounds are deemed to be non-interactive. Additivity results in a linear dose-response behavior, while synergy or antagonism is indicated by non-linear changes in dose-response behavior.[10] While checkerboard assays are most often employed to study combinations of pure compounds, they have also been employed using fractions to study synergy in botanical mixtures.[2,12]

The isobologram approach can be employed as a final validation step to confirm the types of interactions that occur between biologically active compounds.[13] Practically speaking, however, it is not feasible to isolate every constituent from a biologically active natural product mixture and test activity in two-by-two combinations. In cases where biological activity of a mixture may result from the combined effect of multiple compounds, some methodology is needed to help the analyst decide which mixture components to isolate and evaluate.

Several methods have previously been developed to identify synergists from complex natural product mixtures.[14,15] One of these is synergy-directed fractionation[3] in which isolation is guided by measurements of the ability of one compound (or mixture of compounds[1]) to enhance the activity of a known active component of the mixture. With synergy-directed fractionation, it is possible to identify active compounds even if they don't possess activity alone. For example, this approach enabled identification of flavonoids in *Hydrastis canadensis* (goldenseal) that have no inherent antimicrobial activity but enhance the activity of the alkaloid

berberine.[1,3] Building on synergy-directed fractionation, Caesar, et al., developed an approach (called "Simplify") to predict whether features identified in the liquid-chromatography mass spectrometry (LC-MS) datasets for complex mixtures interact synergistically, additively, or antagonistically. Simplify relies on the "activity index," which is a measure of the ratio of the observed activity of a mixture to the activity that would be predicted based on concentration of a known active compound.[2] The Simplify approach was employed to identify sugiol from the medicinal plant *Salvia miltiorrhiza*, and it was shown that sugiol synergistically enhances the antimicrobial activity of the alkaloid cryptotanshinone. Synergy-directed fractionation and Simplify are approaches that can be used to identify constituents in a mixture that enhance the activity (synergistically or additively) of a known active compound. A limitation of these approaches is that they require a-priori knowledge of the identity and concentration of this known active compound.

With the study described here, we set out to develop an approach to identify synergists that would be effective even when none of the active constituents are known. We used untargeted LC-MS metabolomics as a central tool towards this goal. The application of LC-MS metabolomics to identify biologically active natural products relies on the integration of a "chemical" dataset and a "biological" dataset.[4,16,17] The chemical dataset consists of a set of features (ions detected by the mass spectrometer, each described by a characteristic mass to charge ratio, m/z, and retention time) and their associated abundance (peak height or peak area). The biological dataset is a set of measurements that describe how each mixture perturbs a biological system (for example, inhibits cell growth, alters cell morphology, or reduces tumor size in an animal).

Several different data analysis approaches can be used to integrate these chemical and biological datasets. The most intuitive of these is to select the individual features in the chemical dataset one-by-one and compare the abundance profile of each one to the biological activity of the samples. Such comparisons can be accomplished with univariate statistical methods such as Pearson correlation. In scenarios where more than one compound may be responsible for the activity of a natural product mixture, however, multivariate statistical approaches of data analysis are needed. For metabolomics data, multivariate latent-variable regression techniques such as partial-least squares (PLS) regression are particularly appropriate. PLS helps to address the problem of overfitting, which can occur in metabolomics data analysis because the number of observations of biological activity is often small compared with the number of variables (features) in the chemical dataset. Such an "underdetermined" experimental design is more the rule than the exception in natural products metabolomics studies. PLS linearly combines sets of variables that co-vary to create a smaller number of groups of orthogonal variables (latent variables).

Recent studies in natural products have described various workflows to predict active compounds from chemical and biological datasets. These workflows have been referred to with different terms, such as "compound activity mapping,"[18,19] "bioactivity based molecular networking"[20] and "biochemometrics."[4,21] Compound activity mapping and bioactivity based molecular networking successfully identified active compounds from natural product mixtures using Pearson correlations,[18–20] while "biochemometrics" used PLS regression techniques for the identification of single active compounds[4] and mixtures of natural products that act together additively.[22]

A limitation data analysis workflows previously employed for natural products research[4,18–21] is that they have operated under the assumption that the individual mixture components do not interact with each other. Our goal with this study was to develop and test a new "interaction metabolomics" data analysis approach for natural products discovery applicable in scenarios where mixture components interact to achieve biological effects. Towards this goal, we constructed an experimental system for which the observed biological activity was due to the interaction of known synergists. We measured antimicrobial activity against the bacterium *Staphylococcus aureus* in mixtures containing the antimicrobial alkaloid berberine[12,23] and the synergist piperine **(Figure R1)**.[1,24] We also collected mass spectrometry metabolomics data for all of the mixtures, and calculated a set of synthetic features from these data that we refer to as "compound interaction terms" (CIT). Each compound interaction term represents a product of the peak areas of two features detected in the mixtures. Finally, we tested two data analysis workflows, one with these interaction terms included ("interaction metabolomics") and one without the inclusion of the compound interaction terms ("classical metabolomics"). Our ultimate objective was to test whether compound interaction terms in the analysis workflow would enable identification of the synergistic antimicrobial activity of berberine and piperine.

CHAPTER II: EXPERIMENTAL SECTION

General Protocol for Antimicrobial Susceptibility Assay

Antimicrobial susceptibility against *Staphylococcus aureus* (strain SA1199)[25] was evaluated using broth microdilution methods for aerobic bacteria based on the Clinical Laboratory Standards Institute (CLSI) guidelines.[26] Cultures were grown from a single isolated colony of the strain and incubated to log-phase in Müeller-Hinton broth (MHB). The inoculum was diluted into a 96-well plate to achieve a final density of 1.0×10^5 CFU/mL. Samples were introduced in triplicate and diluted in broth with a vehicle of 1%DMSO and 1% glycerol. The negative control was vehicle alone in MHB, and the positive control was the antibiotic levofloxacin. Two sets of wells of the same samples were prepared. The first set (growth wells) contained bacterial inoculum and test compound or control in MHB. The second set of wells (control wells) had the identical composition to the growth wells except that MHB was used in place of bacterial inoculum. After 18 hours of incubation, the optical density of all wells was measured at 600 nm (OD_{600}) using a Synergy H1 microplate reader (Biotek, Winooski, VT, USA). The OD_{600} of each sample was corrected for background absorbance of the sample by subtracting the measured OD_{600} of the control wells from the measured OD_{600} of the growth wells. Antimicrobial activity against *S. aureus* was calculated as percent growth inhibition relative to the vehicle control (**Equation 1**), where vehicle OD_{600} is the background corrected OD_{600}¬ value for the vehicle (1% DMSO, 1% glycerol) in MHB, and sample OD_{600} is the background corrected OD_{600}¬ value for the treatment or control in MHB.

$$\%\ \text{inhibition} = \frac{\text{Vehicle OD}_{600} - \text{Sample OD}_{600}}{\text{Vehicle OD}_{600}} \times 100 \qquad \textbf{(Equation 1)}$$

Synergy Evaluation

To test the combination effects between the model compounds, a broth microdilution checkerboard assay was employed. The antimicrobial berberine (> 98%, Sigma-Aldrich) and the efflux pump inhibitor piperine (>97%, Sigma-Aldrich)[1,24] were tested in combination at concentration ranges of 4.7 to 300 µg/mL and 0.78 to 50 µg/mL respectively. The vehicle used was 1% DMSO and 1% glycerol in MHB. The net fractional inhibitory concentration (∑FIC) was calculated by **Equation 2**, where $MIC_{berberine}$ is equal to the minimum inhibitory concentration of berberine against *S. aureus* alone, $MIC_{piperine}$ is equal to the minimum inhibitory concentration of piperine alone, $MIC_{berberine+piperine}$ is the minimum inhibitory concentration of berberine at a given concentration of piperine, and $MIC_{piperine+berberine}$ is equal to the minimum inhibitory concentration of piperine at a given concentration of berberine.

$$\sum \mathrm{FIC} = \mathrm{FIC}_{\mathrm{berberine}} + \mathrm{FIC}_{\mathrm{piperine}} \qquad \textbf{(Equation 2)}$$

where:

$$\mathrm{FIC}_{\mathrm{berberine}} = \frac{\mathrm{MIC}_{\mathrm{berberine+piperine}}}{\mathrm{MIC}_{\mathrm{berberine}}} \text{ and } \mathrm{FIC}_{\mathrm{piperine}} = \frac{\mathrm{MIC}_{\mathrm{piperine+berberine}}}{\mathrm{MIC}_{\mathrm{piperine}}}$$

For the purposes of this study, combination effects were evaluated based on FIC indices as described previously by Caesar, et al., where $\Sigma FIC \leq 0.5$ is for synergistic effects, ΣFIC between 0.5 and 1.0 is for additive effects, and $\Sigma FIC \geq 4.0$ is for antagonistic effects.[2,27,28]

Preparation of the Simulated Extract

A series of 42 purified natural product compounds for possible inclusion in the simulated extract were obtained from commercial sources (**Table 1**). Each compound was tested for antimicrobial activity against *Staphylococcus aureus* strain SA1199[29] at a concentration of 256 µg/mL. Compounds that demonstrated ≥ 20% inhibition were rejected from the sample set. The remaining compounds were retested at 100 µM in three separate conditions, alone, in

combination with 32 µg/mL berberine (95 µM), or in combination with 32 µg/mL piperine (112 µM) (**Table 2**). Any compound that inhibited bacterial growth by more than 20% under any of these conditions was rejected from the set to be included in the mixture.

A subset of 21 compounds from the original 42 compounds fit the selection criteria of showing ≤ 20% inhibition of *S. aureus* alone or in combination with berberine or piperine. The simulated extract (total mass 1288.8 mg) was prepared from these compounds as follows: naringin (8.2%), betulinic acid (0.2%), atropine (9.7%), amygdalin (12.3%), caffeine (2.9%), chlorogenic acid (2.0%), 3,4-dihydroxybenzaldehyde (2.4%), tropine (9.6%), p-octopamine (2.9%), boldine (10.3%), anisodamine (1.4%), quinine (7.5%), dehydroevodiamine (0.18%), apocynin (7.3%), vanillin (3.0%), ferulic acid (9.5%), vanillic acid (4.6%), syringic acid (4.9%), theobromine (0.5%), stigmasterol (0.6%) and β-sitosterol (0.2%).

Table 1. Antimicrobial screening of 42 natural products at 256 µg/mL for *S. aureus* susceptibility testing

Compound Name	Percent Inhibition	Compound Name	Percent Inhibition
naringin (Sigma, >95%)	1.5 ± 4.1	p-octopamine (Cayman, >98%)	5.6 ± 4.9
beta-sitosterol (Sigma, >96%)	-9.56 ± 0.67	18β-glycyrrhetinic acid (Cayman, >98%)	57 ± 13
betulinic acid (Sigma, 90%)	8 ± 16	loganin (Cayman, >98%)	-7.1 ± 9.6
stigmasterol (Sigma, >95%)	-9.3 ± 2.1	4-hydroxycoumarin (Cayman, >98%)	84.0 ± 2.3
atropine (Sigma, >99%)	2.29 ± 0.30	7-methoxyflavone (Cayman, >98%)	20 ± 30
capsaicin (Sigma, >50%)	18.5 ± 7.1	boldine (Cayman, ≥ 95%)	0.7 ± 3.4
amygdalin (Sigma, >97%)	10.7 ± 1.9	anisodamine (Cayman, > 98%)	4.0 ± 1.6
chrysin (Sigma, >98%)	70 ± 14	harmine (Cayman, >98%)	34 ± 19
quercetin (Sigma, >98%)	94.7 ± 5.2	quinine (Cayman, >95%)	-3.4 ± 1.7
caffeine (Sigma, >99%)	3.0 ± 2.1	ursolic acid (Cayman, >98%)	21.1 ± 7.7
kaempferol (Alfa aesar, >98%)	50.9 ± 6.8	dehydroevodiamine (Cayman, >98%)	16.7 ± 1.6
myricetin (TCI, >97%)	98.41 ± 0.69	2-hydroxyanthraquinone (Cayman, >98%)	69 ± 21
chlorogenic acid (Alfa aesar, 98.2%)	16.1 ± 1.9	apocynin (Cayman, >98%)	6.1 ± 1.4
rutin (Arcos Organics, 97%)	38.2 ± 1.3	etoposide (Cayman, >98%)	99.2 ± 4.3
isorhynchophylline (Cayman, >95%)	-12.11 ± 0.74	vanillin (Alfa aesar, >99%)	-0.7 ± 3.6
3,4-dihydroxybenzaldehyde (Cayman, >98%)	11.3 ± 4.8	coumarin (TCI, >99%)	67.8 ± 3.0
palmatine (Cayman, >98%)	99.04 ± 0.07	salicylic acid (SCBT, 99%)	41.1 ± 3.6
tropine (Cayman, >95%)	-0.1 ± 2.3	ferulic acid (Cayman, >98%)	8.0 ± 2.6
chrysosplenetin (Cayman, >98%)	-12.5 ± 1.9	vanillic acid (SCBT, 97%)	20.3 ± 3.5
naringenin (Cayman, >98%)	94.7 ± 4.0	syringic acid (Cayman, 98%)	8.9 ± 1.5
berberine (Sigma, >98%)	99.84 ± 0.12	theobromine (Cayman, >98%)	5.9 ± 4.2

Table 2. Antimicrobial activity of pre-selected compounds at 100 µM and in combination with berberine (32 µg/mL) and with piperine (32 µg/mL)

Compound Name	Percent Inhibition at 100 µM		
	Compound only	With berberine at 32 µg/mL	With piperine at 32 µg/mL
naringin	3.0 ± 1.4	35.1 ± 1.6	13.1 ± 1.6
betulinic acid	13 ± 14	27 ± 13	11 ± 14
atropine	4.48 ± 0.70	41.93 ± 0.77	18.2 ± 5.1
capsaicin	-16.8 ± 2.0	99.16 ± 0.13	16.4 ± 1.7
amygdalin	1.0 ± 2.9	32.4 ± 3.1	15.29 ± 0.95
caffeine	4.8 ± 1.6	35.29 ± 0.57	17.5 ± 1.3
chlorogenic acid	6.46 ± 0.78	36.1 ± 1.8	20.8 ± 4.2
3,4-dihydroxybenzaldehyde	2.3 ± 2.1	29.1 ± 3.2	18.2 ± 4.3
tropine	0.4 ± 3.1	31.6 ± 3.9	14.5 ± 2.4
p-octopamine	4.3 ± 1.1	35.10 ± 0.35	18.1 ± 1.1
boldine	4.5 ± 1.6	34.3 ± 2.6	16.6 ± 1.1
anisodamine	0.8 ± 3.4	29.7 ± 3.2	15.1 ± 2.9
quinine	-1.6 ± 3.7	29.1 ± 1.6	11.2 ± 2.8
dehydroevodiamine	6.3 ± 1.5	32.0 ± 2.4	19.1 ± 1.1
apocynin	5.4 ± 1.8	28.31 ± 0.92	16.5 ± 1.4
vanillin	2.5 ± 2.6	24.6 ± 6.5	17.1 ± 1.8
ferulic acid	-0.9 ± 7.3	29.2 ± 2.9	16.5 ± 3.0
vanillic acid	5.4 ± 1.2	32.7 ± 1.8	14.3 ± 1.8
syringic acid	4.38 ± 0.65	29.84 ± 0.59	15.7 ± 2.0
theobromine	0.3 ± 4.0	25.8 ± 6.6	15.4 ± 2.7
berberine	2.0 ± 7.2	48.6 ± 1.9	99.35 ± 0.11
piperine	-0.9 ± 1.6	99.12 ± 0.18	23.1 ± 3.2

Chromatographic Separation of the Simulated Extract

Solvents used in chromatographic separation were ACS grade (Fisher Scientific). The simulated extract was dissolved completely in methanol and fractionated by normal-phase flash column chromatography on a CombiFlash RF system with a 40 g silica gel column. Gradient elution using hexane, chloroform, and methanol was employed for 84 minutes at a flow rate of 40 mL/min and yielded 150 eluates. The eluates were pooled in sets of 15 tubes to make 10 fractions. The first pooled fraction was not used because it contained an insufficient quantity of material. The rest of the pooled fractions were labeled 01-09 and were used as background matrices for preparation of the spiked fractions.

Antimicrobial Evaluation of Spiked Fractions

Two sets of mixtures ("spiked fractions") were prepared by spiking the inactive background matrices with berberine (M01-M08, **Table 3**) or with berberine and piperine (M09-M17, **Table 4**). Pooled fractions 01-08 were used as background matrices for M01-M08 (**Table 3**), and pooled fractions 01-09 were used to create M09-M17 (**Table 4**). Thus, pooled fractions 01-08 were used twice as background matrices, once for the berberine mixtures and once for the berberine-piperine mixtures, while pooled fraction 09 was used only once for one berberine-piperine mixture. (Fewer berberine mixtures were required for testing than berberine-piperine mixtures.) To create the mixtures with final assay concentrations shown in **Table 3** and **Table 4**, stock solutions of berberine and piperine were prepared at a concentration of 10 mg/mL in DMSO-glycerol (1:1) and combined with stock solutions of the pooled fractions. The concentrations of berberine used in M01-M08 (**Table 3**) was slightly higher than that used in M09-M17 (**Table 4**) to avoid saturating the antimicrobial response due to the impact of added piperine. Three separate antimicrobial broth dilution assays were performed on the mixtures

using the "General Protocol for Antimicrobial Susceptibility Testing," one with the nine pooled fractions (background matrices) alone, one with the pooled fractions spiked with berberine, and one with the pooled fractions spiked with berberine and piperine.

LC-MS Analysis of Spiked Fractions

Each spiked fraction was diluted 100-fold from assay concentration and analyzed in triplicate using a Thermo Fisher Q Exactive™ Plus mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with an electrospray ionization (ESI) source coupled with a Waters Acquity ultraperformance liquid chromatograph (Waters Corporation, Milford, MA). A 3 µL volume of each sample was injected and eluted through a reversed-phase column (BEH C18, 1.7 µm, 2.1 x 50 mm, Waters Corporation) using a binary solvent system consisting of water with 0.1% formic acid (solvent A) and acetonitrile with 0.1% formic acid (solvent B). The 10-min. gradient elution started with 10%B for 0.5 min. then increased to 100%B for 8 min. and finally re-established to starting conditions in the last 1.5 min. Analysis was conducted in full scan acquisition, collecting profile data in switching positive and negative polarity. The scan range was from 120 to 1500 m/z with a scan time of 200 ms. The following mass spectrometer parameters were used: the AGC target was at 1×10^6 with a capillary voltage and temperature at -0.7 V and 310 °C respectively; the S-lens RF level was 80.00, spray voltage was 3.7 kV, and the sheath and auxiliary gas flows were 50.15 and 15.16, respectively. The full MS dataset is uploaded and accessible as MassIVE dataset MSV000089598 (ftp://MSV000089598@massive.ucsd.edu).

Metabolomics Data Peak Picking and Data Filtering

Peak Picking. The two different LC-MS datasets (M01-M08 and M09-M17) were analyzed separately. LC-MS raw files were imported to MZmine 2.53[30] for peak picking.

Methods to create a list of features include mass detection, chromatogram building, chromatogram deconvolution, deisotoping, feature alignment, gap-filling, duplicate filter, and peak filter and the parameters used are included in **Table S1**. The final feature lists were imported to MS Excel for further treatment.

Data Filtering. LC-MS data were first filtered to remove background noise, which may be due to small solvent contaminants or electronic noise in the system. The first blank filter aims to remove these signals, the relative standard deviation (RSD) of a feature across all samples, including the blanks, were calculated. All features with RSD across samples and blanks that were less than 30% were removed from the dataset. The next blank filter was to remove features with higher signals in the blanks as compared to the samples. This was done by calculating the percent ratio of the mean of the blanks to the mean of the samples. All features with a percent ratio that were higher than 80% were removed. The datasets were then filtered to remove poor quality features based on their having greater than a relative standard deviation (RSD) cutoff of 35% RSD. The RSD was calculated from the feature peak areas of triplicate LC-MS analyses of the same sample. The full, RSD and blank filtered feature lists are available as supporting information (https://doi.org/10.5281/zenodo.6612585). Peak areas shown are the average peak area across the three replicate injections. The second filtering step removed mass spectral features that did not vary in intensity (peak area) across the spiked fractions (mixtures), choosing an empirically selected cutoff value of $\leq 0.01\%$ of the variance for the feature with highest variance for M01-M08 and $\leq 0.1\%$ of the variance of the features with highest variance for M09-M17. The reduced dataset is available as Supporting information (https://doi.org/10.5281/zenodo.6612585).

Calculation of the Compound Interaction Terms

The compound interaction terms (CIT) were calculated with **Equation 3**, where I_{F_i} represents the intensity of feature *i*, $I_{F_{(i+1)}}$ represents the intensity of feature (*i* + 1), and $CIT_{F_iF_{(i+1)}}$ represents the compound interaction term for features *i* and (*i*+1).

$$CIT_{F_iF_{(i+1)}} = I_{F_i} \times I_{F_{(i+1)}} \quad \textbf{(Equation 3)}$$

For a given dataset, the total number of non-redundant compound interaction terms (N_{CIT}) is determined by **Equation 4**, where m is equal to the total number of features (ions) detected across all samples.

$$N_{CIT} = \frac{m(m+1)}{2} \quad \textbf{(Equation 4)}$$

Data Standardization

Highly abundant features will have a large variance that will dominate the first PLS components. Thus, they may mask other features, leading to misinterpretation of the data. This problem is magnified when compound interaction terms are calculated because the magnitude of the compound interaction terms is large compared to the magnitude of the individual features. To overcome this scaling problem, LC-MS peak areas or interaction terms were standardized to unit variance.[31] To calculate the standardized abundance of each feature (I_{std,F_i}), the intensity of each feature (I_{F_i}) in each mixture was divided by the standard deviation (s) of the peak area of that feature across all samples (mixtures) (**Equation 5**).

$$I_{std,F_i} = \frac{I_{F_i}}{s} \quad \textbf{(Equation 5)}$$

The standardized compound interaction term intensities were calculated in the same fashion as the standardized feature intensities, using the compound interaction term (CIT) values in place of the feature peak areas.

Statistical Analysis: Partial-Least Squares Regression and Calculation of Selectivity Ratios

The pre-processed datasets were imported to Sirius 11.5 (Pattern Recognition Systems AS, Bergen, Norway) for statistical analysis. The data were modeled using partial least-squares regression followed by target projection (TP) to obtain selectivity ratios that connect biological activity to the mass spectral variables.[4,16,17,32] To ensure the reliability of the model and good line fitting between the predicted and measured response variables, the root-mean-squared error of prediction (RMSEP) must be calculated through validation based on repeated Monte Carlo resampling.[33–35] The dataset was validated leaving out one object with 100 repetitions and significance level of 0.5 to determine the number of PLS components to use for modeling. Selectivity ratios of mass spectral variables (or LC-MS features) correlating with biological activity were calculated to identify compounds of interest.[1,2,4,16,17,32,36]

CHAPTER III: RESULTS AND DISCUSSION

For this study, we created an experimental design where all five of the requirements to observe synergy (see *Introduction*) were satisfied. Antimicrobial activity against *Staphylococcus aureus* was selected as the biological effect to be measured, and the number of mixtures necessary to distinguish a synergistic effect was determined using a modified factorial design.[37] We selected a known antimicrobial (berberine, **Figure R1**[12]) and a known synergist (piperine, **Figure R1**[1,24]), both of which are detectable by LC-MS. We measured the dose-response behavior of these compounds alone and in combination (**Figure 1**) and prepared a series of mixtures at a range of concentrations where direct antimicrobial activity of berberine (**Table 3**) or synergistic antimicrobial activity of berberine and piperine (**Table 4**) would be observed. To create an inactive matrix for the mixtures, we prepared a simulated extract by mixing inactive compounds and separating them using flash chromatography. The antimicrobial activity of the pooled fractions from the simulated extract were shown to be inactive against *S. aureus* (**Figure 2**) and served as the background matrix for the berberine or berberine-piperine mixtures (spiked fractions). Finally, we measured the antimicrobial activity of each of the spiked fractions and subjected each to analysis with LC-MS. The result of these experiments was a dataset that we could use to develop and validate the interaction metabolomics approach to identify synergists. Notably, this dataset, which is freely available (https://doi.org/10.5281/zenodo.6612585) could also be employed by scientists seeking to benchmark other methodologies for identifying synergists.

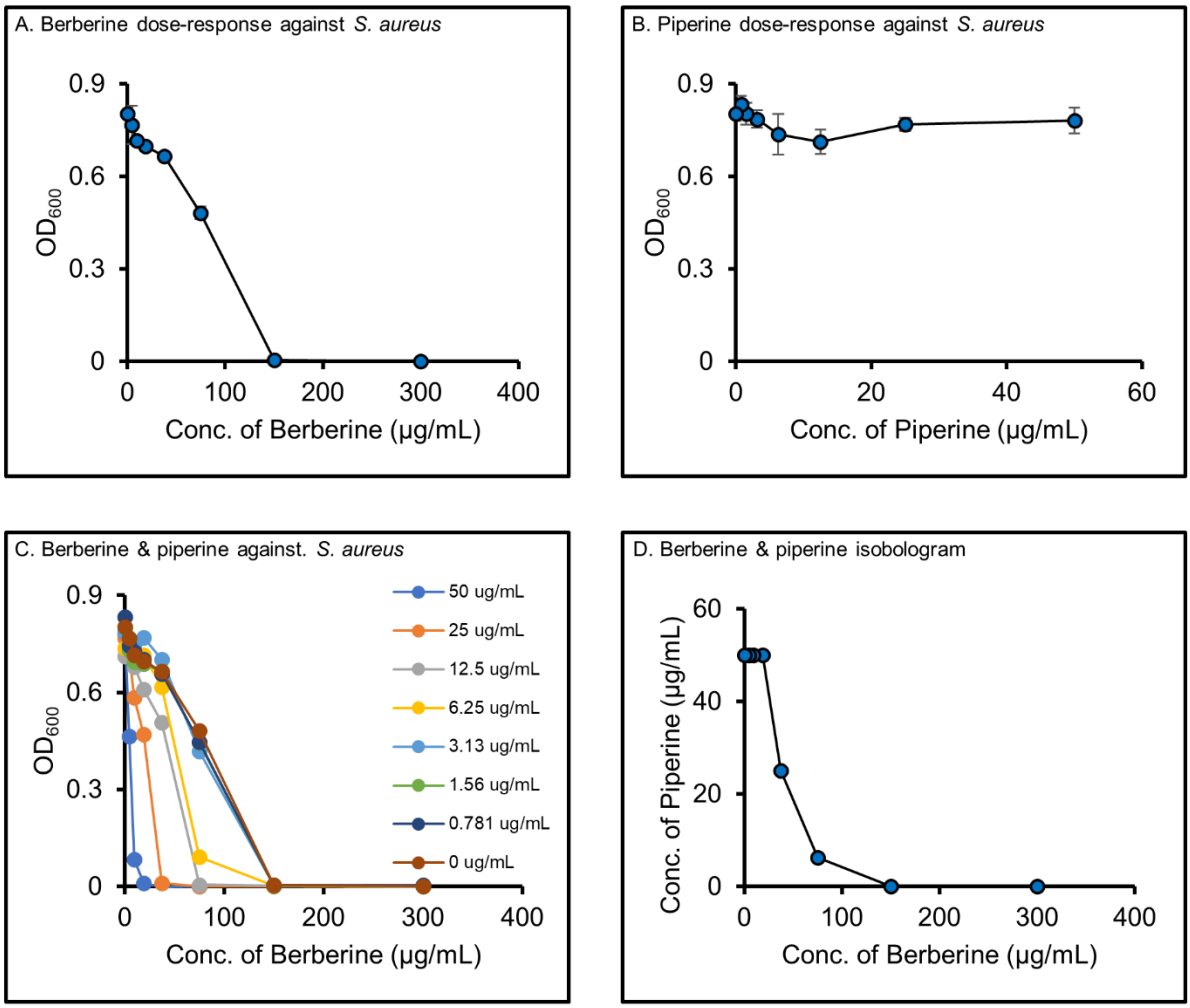
Antimicrobial Activity of Berberine and Piperine is an Effective Model for Synergy

Berberine and piperine were selected for these studies as an antimicrobial and synergist, respectively. Berberine has been reported to inhibit the growth of *S. aureus*,[3,12,23] while piperine

had been reported to act as an efflux pump inhibitor.[1,24] Berberine is a substrate to the NorA efflux pump in *S. aureus*, and piperine inhibits bacterial efflux, enhancing the antimicrobial activity of berberine without possessing any direct antimicrobial activity.[3,5,12]

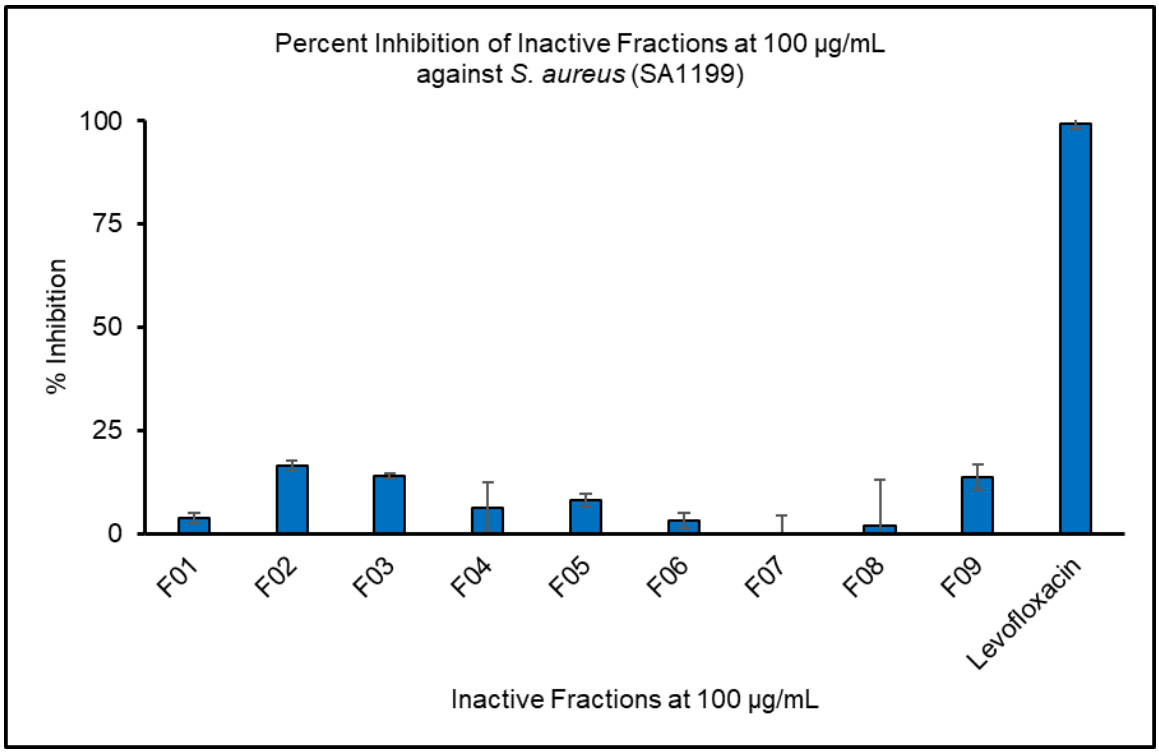
To confirm synergy between berberine and piperine, a checkerboard assay was conducted (**Figure 1**). Results show that berberine alone inhibits the growth of *S. aureus* with an MIC of 150 µg/mL (446 µM) as previously reported[23] (**Figure 1A**), while piperine alone does not show measurable antimicrobial activity (**Figure 1B**). As berberine is combined with increasing concentrations of piperine (**Figure 1C**), its MIC shifts to as low as 9.38 µg/mL (27.9 µM). An isobologram was plotted (**Figure 1D**) and using **Equation 2**, the net fractional inhibitory concentration index of berberine and piperine (∑FIC) was calculated to be 0.19, which is ≤ 0.50, demonstrating synergy. Therefore, berberine and piperine possess synergistic antimicrobial activity against *S. aureus* under the conditions used for this study.

Figure 1. Checkerboard assay results of berberine and piperine combinations against *S. aureus*



Notes. Figures 1A and 1B show the dose-response curves of berberine and piperine against *S. aureus*, respectively. Figure 1C shows a dose-response curve of berberine combined with different concentrations of piperine (shown in different colors). Without added piperine, the MIC of berberine is 150 µg/mL. The addition of piperine reduces the MIC of berberine. For example, the MIC of berberine is 9.38 µg/mL in the presence of 50 µg/mL piperine. Figure 1D shows a hyperbolic isobologram indicating synergy. Additionally, the ∑FIC value (**Equation 2**) of berberine in the presence of 50 µg/mL piperine is 0.19, signifying synergy.

Figure 2. Antimicrobial activity against *Staphylococcus aureus* of the fractions without berberine or piperine



Notes. Error bars represent the standard deviation of triplicate % inhibition values. Percent inhibition of bacterial growth is expressed relative to the vehicle control (1% DMSO, 1% glycerol). The antibiotic levofloxacin (10 µg/mL) served as the positive control.

Characteristics of the Biological and Chemical Datasets Used to Model Synergy

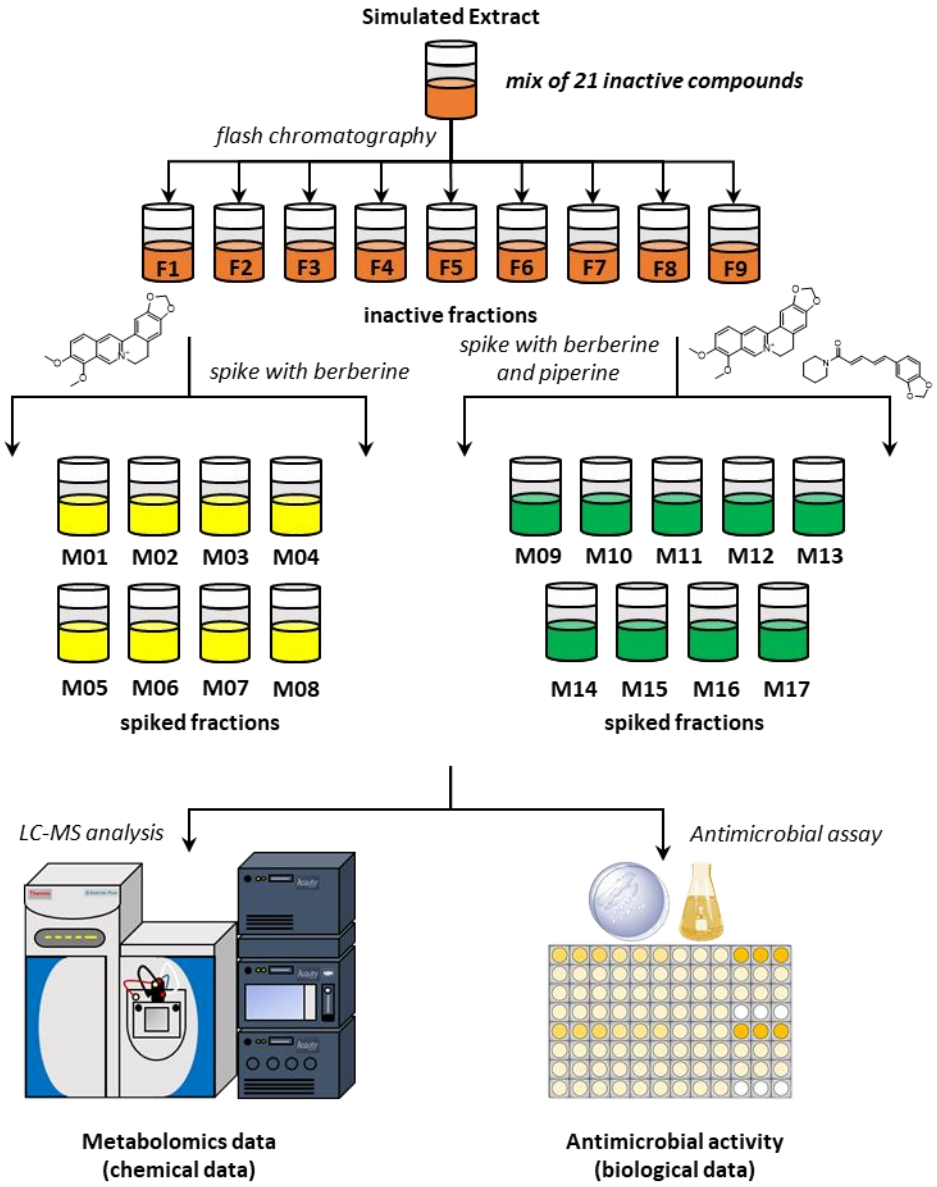
We employed a modified factorial design[37] that enabled a demonstration of the interaction effects between berberine and piperine using a smaller number of mixtures (total of 9) than is typically employed to collect an isobologram.[13] To conduct this experiment, we prepared nine mixtures with known concentrations of berberine and piperine, each with an inactive background matrix. The inactive matrices were created using flash chromatographic separation of a simulated natural product extract (**Figure 3**). Thus, each fraction had varying levels of the inactive constituents similar to what would be obtained by a typical natural products isolation experiment. We controlled the levels of berberine and piperine by spiking them into the

mixtures after chromatographic separation because we wished to create a test system where synergy would certainly occur. In natural products drug discovery project where the identities of synergists are not known, it is possible that synergy could be missed because synergists would be separated from each other in the fractionation process. One possible strategy to prevent this from occurring would be to accomplish "poor" chromatography, with broad peaks, and to generate many fractions. It would also be possible to increase the likelihood that synergists co-occur by intentionally recombining fractions after chromatographic separation, as has been done previously.[12,22,38] The optimization of methods for fractionation and recombination to observe synergy would be a worthy topic of future research. Our goal with this study was to test methods of assigning active constituents when synergists are present.

To prepare the simulated extract, we tested the antimicrobial activity of a series of 42 commercially available natural products (**Table 1**) to select those that fulfilled the following two selection criteria: (1) did not inhibit *S. aureus* growth by >20% (**Equation 1**) at assay concentration of 100 µM and (2) did not enhance the antimicrobial activity of berberine or piperine. A subset (22) of the compounds tested fit criterion 1 (**Table 1**). When these compounds were tested in combination with berberine (32 µg/mL or 95 µM) or piperine (32 µg/mL or 112 µM), none enhanced the activity of piperine and only capsaicin enhanced the antimicrobial activity of berberine (**Table 2**). The observed enhancement of berberine activity by capsaicin was consistent with a previous report of capsaicin being an inhibitor of the NorA efflux pump of *S. aureus*.[39] Therefore, capsaicin was not included in the simulated mixture reducing the number of compounds to 21. Finally, the components were all analyzed with LC-MS and all except stigmasterol and β-sitosterol were detected in extracted ion chromatograms and by peak picking above the threshold of 1×10^5 using MZmine 2.53[30] (**Table S1**).

Flash chromatography was employed to separate the simulated extract into ten pooled fractions. From these, the nine pooled fractions that yielded sufficient material for the experiments were spiked to create a series of fractions spiked with berberine alone (M01-M08) (**Table 3**) or with berberine and piperine (M09-M17) (**Table 4**). Different concentrations of berberine were used between the two sets of mixtures because berberine antimicrobial activity was expected to saturate at higher concentrations in the mixtures without piperine. The background matrices used to create the spiked fractions were prepared with a final assay concentration of 100 µg/mL (expressed as mass of dried material per mL assay volume). As in a realistic natural products isolation experiment, the concentrations of individual background matrix components in these mixtures were not known.

Figure 3. Experimental workflow for preparation and analysis of spiked fractions



Notes. A simulated extract was prepared by mixing 21 natural products that did not demonstrate antimicrobial activity against *S. aureus*, alone or in combination with berberine and piperine. This mixture was fractionated and the pooled fractions were used as background matrices to create spiked fractions containing known amounts of berberine (antimicrobial) and piperine (synergist). Untargeted metabolomics data were collected with ultraperformance liquid chromatography coupled to mass spectrometry (LC-MS) to obtain a metabolomics dataset. Antimicrobial activity was evaluated for all fractions against *Staphylococcus aureus* to obtain a biological dataset.

Table 3. Composition and biological activity of berberine spiked fractions without piperine

Sample	Berberine (µg/mL)	Piperine (µg/mL)	Fraction No.[a]	% inhibition (± s.d.)[b]
M01	0	0	F01	5.4 (±8.9)
M02	100	0	F02	97.67 (±0.34)
M03	75	0	F03	97.50 (±0.64)
M04	64	0	F04	88.4 (±8.8)
M05	50	0	F05	29.0 (±4.5)
M06	32	0	F06	20.47 (±0.13)
M07	25	0	F07	16.9 (±2.1)
M08	16	0	F08	13.7 (±1.8)
Berberine (at 100 µg/mL)[c]				95.1 (±2.7)
Levofloxacin (positive control)[c]				99.61 (±0.22)

Notes. Concentrations indicate the assay concentrations used to evaluate % inhibition against *Staphylococcus aureus*. Prior to analysis by LC-MS, samples were diluted 100-fold from the concentrations shown below to avoid saturation of instrument response.

[a]Fraction indicates the pooled mixture of inactive compounds generated by flash chromatography separation of the simulated extract. Concentration is expressed as mass of the total mixture per well volume and does not indicate concentration of individual mixture components. Each mixture used a different fraction containing some subset of the compounds from the simulated extract. Each fraction was added at a concentration of 100 µg/mL, expressed as mass dried fraction per well volume in the antimicrobial assay.

[b]Antimicrobial activity is expressed as % inhibition relative to vehicle control ± standard deviation across triplicate wells.
[c]Levofloxacin and berberine were used as positive controls at 10 µg/mL and 100 µg/mL respectively.

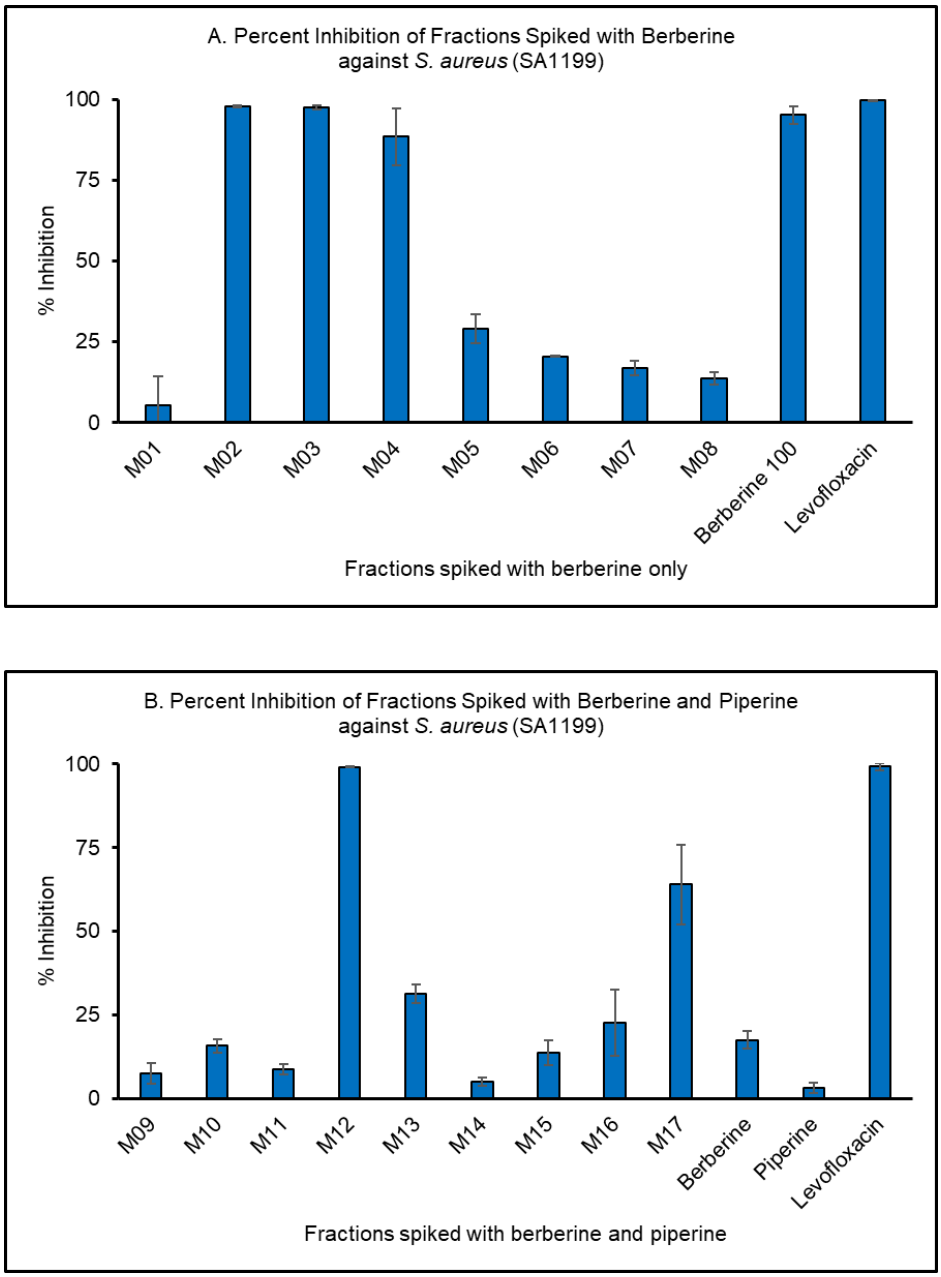
Table 4. Composition and biological activity of berberine and piperine spiked fractions

Sample	Berberine (µg/mL)	Piperine (µg/mL)	Fraction No.[a]	% inhibition (± s.d.)[b]
M09	0	0	F01	7.4 (±3.1)
M10	32	0	F02	15.8 (±2.0)
M11	0	32	F03	8.7 (±1.6)
M12	32	32	F04	99.12 (±0.15)
M13	16	16	F05	31.2 (±2.8)
M14	8	8	F06	5.0 (±1.2)
M15	24	8	F07	13.7 (±3.7)
M16	8	24	F08	22.7 (±9.8)
M17	24	24	F09	64 (±12)
Berberine (at 32 µg/mL)[d]				17.5 (±2.7)
Piperine (at 32 µg/mL)[d]				3.2 (±1.4)
Levofloxacin (positive control)[c]				99.3 (±1.1)

[a]The same background matrices (fractions) used for mixtures 01-08 were used to prepare mixtures 09-16 here, plus another fraction for mixture 17. All fractions were tested at an assay concentration of 100 µg/mL.
[b]Antimicrobial activity is expressed as % inhibition of *S. aureus* growth relative to vehicle control ± standard deviation among triplicate wells.
[c]Levofloxacin was used as the positive control at 10 µg/mL.
[d]Berberine and piperine at 32 µg/mL were added as controls.

To obtain the "biological dataset," antimicrobial activity was measured for the pooled fractions alone (**Figure 2**) and for the fractions spiked with berberine (**Figure 4A**) or berberine and piperine (**Figure 4B**). As expected, the pooled fractions alone (without berberine or piperine) demonstrated less than 20% growth inhibition of *S. aureus* (**Figure 2**). Spiking the fractions with berberine alone (M01-M08) caused a dose-dependent inhibition of bacterial growth (**Figure 4A**). The mixtures spiked with berberine and piperine (**Figure 4B**) also suppressed bacterial growth, with the highest activity observed for mixture twelve (M12). Synergistic antimicrobial activity was observed for several of the mixtures (**Figure 4B**).

Figure 4. Antimicrobial activity against *Staphylococcus aureus* of the spiked fractions, with berberine only (A) and with berberine and piperine (B)



Notes. Error bars represent the standard deviation of triplicate % inhibition values. Percent inhibition of bacterial growth is expressed relative to the vehicle control (1% DMSO, 1% glycerol). Levofloxacin (10 µg/mL) served as the positive control as a known antibiotic. In A, berberine at 100 µg/mL was also added as a control. The synergistic enhancement of biological activity by piperine is apparent in **Figure 4B**. For example, mixture ten (M10) contains 32 µg/mL berberine and only exhibits 15.8 ± 2.0% inhibition of *S. aureus* growth. Mixture twelve (M12) contains the same amount of berberine as M10 (32 µg/mL) but also contains 32 µg/mL piperine and demonstrates 99.12 ± 0.15% inhibition. Since piperine demonstrates no antimicrobial activity alone, the enhanced biological activity of M12 as compared to M10 represents synergy. A similar conclusion can be drawn by comparing the biological activity of M15 to M17 (**Figure 4B**).

LC-MS analysis was conducted on all of the spiked fractions to obtain the "metabolomics dataset" (**Figure S1** to **Figure S4**). Berberine was readily detectable in M01-M08 (**Figure S1** and **Figure S2**), while berberine and piperine were detectable in M09-F17 (**Figure S3** and **Figure S4**). Berberine and piperine were identified by their characteristic mass spectral data (**Figure S5**). The mass spectrum for berberine (**Figure S5A**), which has an inherent positive charge, is characterized by a single peak representing the $[M^+]$ ion. Piperine is detected as a series of five features (**Figure S5B**), including the protonated species $[M+H]^+$, sodiated $[M+Na]^+$ species, proton bound dimer, $[2M+H]^+$, sodium bound dimer $[2M+Na]^+$, and a sodiated acetonitrile cluster $[M+ACN+Na]^+$.

Analysis of standards indicated that all of the inactive matrix compounds were detectable by LC-MS except stigmasterol and β-sitosterol (**Table 5**), for a total of 19 detectable compounds. Similarly, and as would be expected, each of the inactive matrix compounds were also detectable in at least one of the spiked fractions (**Table 6**). Of the 19 components detected in the original dataset (**Table 5**), seven (naringin, chlorogenic acid, tropine, p-octopamine, vanillic acid, and theobromine) were removed in the data filtering step that required variation in abundance across the mixtures (**Table 6**). Thus, the final metabolomics datasets included features (**Table S2**) from berberine, piperine, and twelve inactive compounds.

Table 5. Distribution of analytes in the spiked fractions as detected by LC-MS after filtering based on <35% relative standard deviation in peak area across replicate injections

Analyte	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12	M13	M14	M15	M16	M17
naringin								■								■	
betulinic acid											■	■					
atropine	■						■	■	■						■	■	■
amygdalin								■								■	■
caffeine				■	■							■	■				
chlorogenic acid									■							■	■
3,4-dihydroxybenzaldehyde	■				■	■	■		■				■	■	■	■	■
tropine	■								■								■
p-octopamine																	■
boldine	■					■	■		■					■	■	■	■
anisodamine	■								■								■
quinine	■				■	■	■	■	■				■	■	■	■	■
dehydroevodiamine	■						■	■	■						■		■
apocynin		■	■	■						■	■	■	■				
vanillin		■	■							■	■	■	■	■			
ferulic acid					■	■	■							■	■	■	
vanillic acid														■	■	■	
syringic acid					■									■	■	■	
theobromine													■				
β-sitosterol																	
stigmasterol																	
berberine		■	■	■	■	■	■	■		■		■	■	■	■	■	■
piperine											■	■	■	■	■	■	■

Black – analyte identified
White – not detected

Table 6. Distribution of analytes in the simulated fractions after filtering features that do not vary across samples

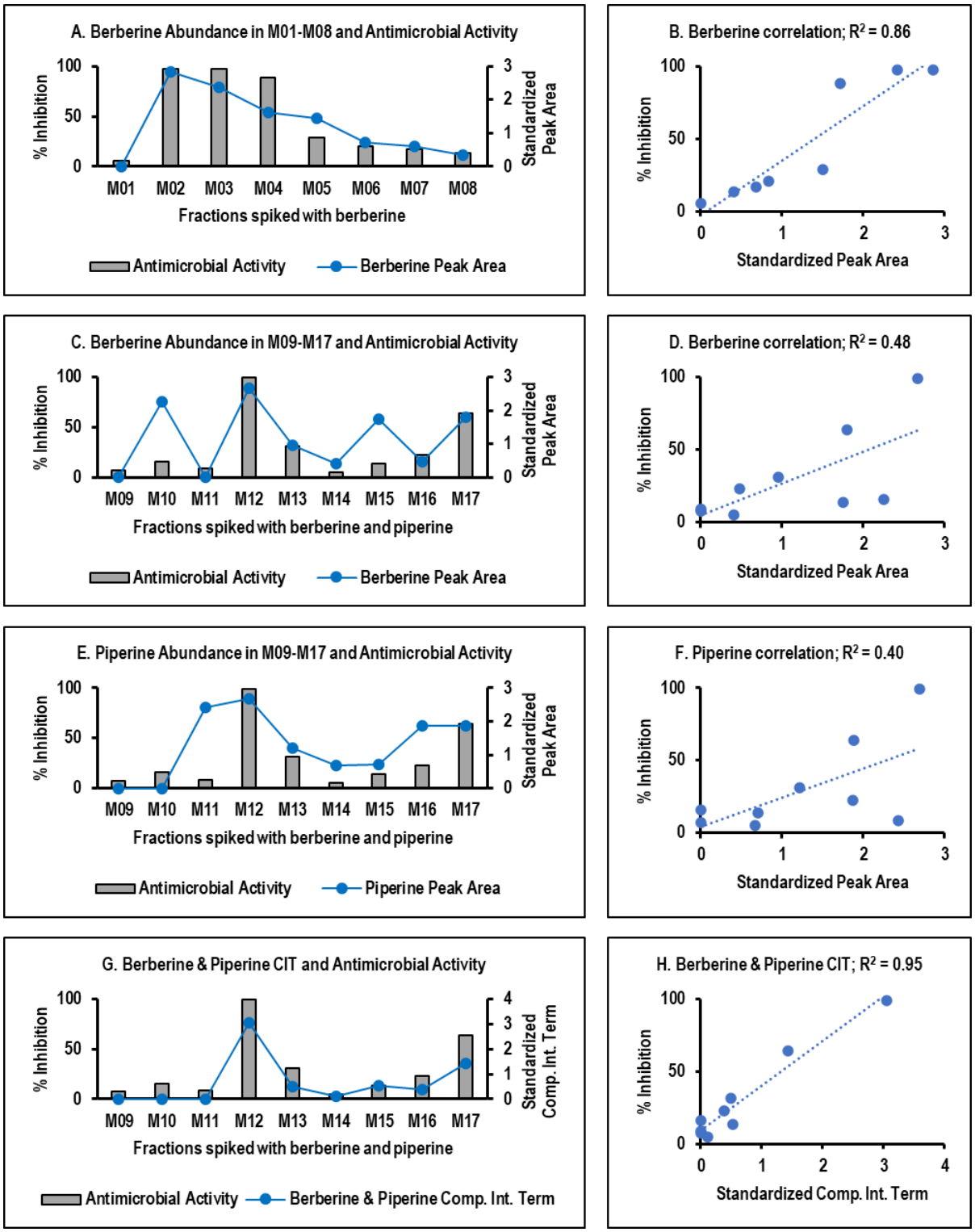
Analyte	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12	M13	M14	M15	M16	M17
naringin								■									
betulinic acid																	
atropine	■						■	■	■						■	■	■
amygdalin								■								■	■
caffeine				■	■							■	■				
chlorogenic acid																	
3,4-dihydroxybenzaldehyde	■				■	■	■		■				■	■	■	■	■
tropine																	
p-octopamine																	
boldine	■					■	■		■					■	■	■	■
anisodamine	■								■								■
quinine	■				■	■	■	■	■				■	■	■	■	■
dehydroevodiamine	■						■	■	■								
apocynin		■	■	■						■	■	■	■				
vanillin										■	■	■	■	■			
ferulic acid					■	■	■							■	■	■	
vanillic acid																	
syringic acid					■												
theobromine																	
β-sitosterol																	
stigmasterol																	
berberine		■	■	■	■	■	■	■		■		■	■	■	■	■	■
piperine											■	■	■	■	■	■	■

[a]Black – analyte identified
[b]White – not detected or filtered out
[c]% Variance cut-off: M01-M08 = 0.1%; M09-M14 = 0.01%

Conceptual Demonstration of the Compound Interaction Term

To demonstrate the concept of the compound-interaction term (**Equation 3 and Equation 4**), we examined the data obtained from the chemical and biological analysis of the spiked fractions by selecting major features associated with berberine ($[M^+]$) and piperine ($[M+H]^+$) and comparing their standardized abundance (**Equation 5**) with the measured biological activity of the fractions (**Figure 5**). For the fractions spiked with just berberine (M01-M08, **Figure 5A**), biological activity is correlated with the peak area of berberine. For the fractions that contain both berberine and piperine (M09-M17), neither the peak area of the berberine feature (**Figure 5C**, **5D**) nor the peak area of the piperine feature (**Figure 5E, 5F**) is strongly correlated with biological activity. This is expected given that the antimicrobial activity of the mixtures results from the combined (synergistic) activity of berberine and piperine. To obtain a value that correlated with activity in the berberine-piperine mixtures, a compound interaction term (**Equation 3**) was obtained by multiplying the peak area of one berberine feature ($[M^+]$) by the peak area for one piperine feature ($[M+H]^+$). This compound interaction term tracks closely with antimicrobial activity (**Figure 5G**) and the relationship is linear (R^2 value of 0.95) (**Figure 5H**).

Figure 5. Comparison of biological and chemical data demonstrate the utility of the compound interaction term

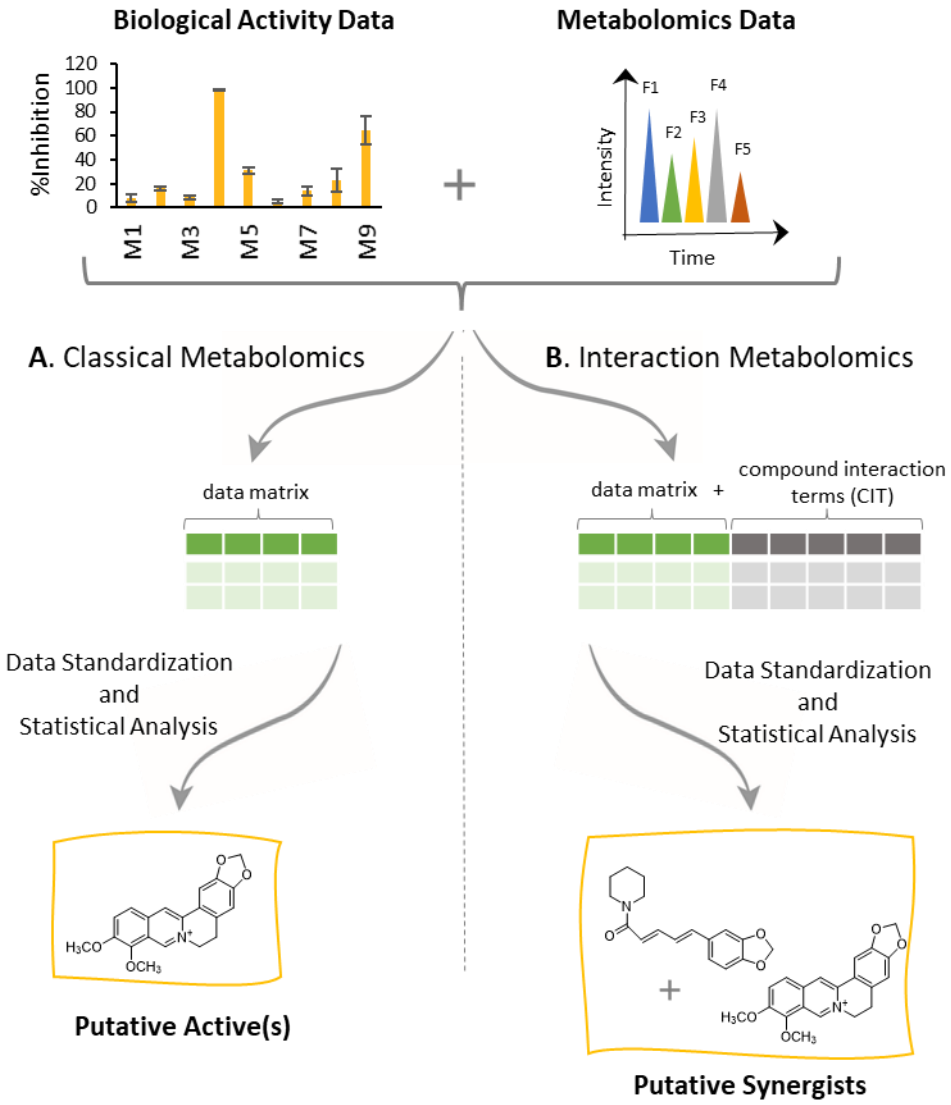


Notes. The antimicrobial activity (% inhibition of *S. aureus* growth) was strongly correlated with standardized abundance for the fractions spiked with berberine (M01-M08, panel A and B). For the mixtures spiked with berberine and piperine (M09-M17), neither berberine (C) nor piperine (E) abundance tracks with biological activity. This lack of correlation is shown with poor linearity of regression plots for % inhibition versus peak area of berberine (D) and piperine (F). The compound interaction term, obtained by multiplying the peak area for piperine with the peak area for berberine (**Equation 3**), tracks with biological activity for the berberine-piperine mixtures (G). The linear relationship between % inhibition and compound interaction term is demonstrated in panel H. To generate these plots, piperine abundance was measured as the peak area of the selected ion for the $[M+H]^+$ ion of piperine detected at m/z 286.1426 while peak area of berberine was measured as the peak are of the $[M]^+$ ion detected at m/z 336.1217. The peak areas for these ions were normalized as shown by **Equation 5**.

Comparison of Classical and Interaction Metabolomics Workflows

While a univariate approach, as depicted in **Figure 5**, is useful for demonstrating the compound interaction term idea, a multivariate statistical approach is more appropriate when untargeted metabolomics datasets containing many features are analyzed. Here we analyzed the data in a "classical metabolomics" workflow (**Figure 6A**) and an "interaction metabolomics" workflow (**Figure 6B**). The difference between the two workflows was in the data matrix (**Figure 7**) used for the analysis. The classical workflow was conducted on a matrix containing biological activities and standardized feature intensities for each of the mixtures. In the workflow allowing interactions, the matrix was expanded to include standardized compound interaction terms (**Equation 3**) for each feature pair in the dataset. Both datasets used in these calculations are freely available (https://doi.org/10.5281/zenodo.6612585).

Figure 6. Two possible metabolomics workflows for data analysis, (A) classical metabolomics and (B) interaction metabolomics



Notes. The two workflows both start with a biological dataset and a metabolomics dataset (in this case LC-MS data obtained by analysis each of the mixtures individually). The values from these biological and metabolomics datasets are compiled in one of two data matrices **(Figure 7)**. The data matrix for the interaction workflow differs by the inclusion of an additional compound interaction terms (CIT). The values in the data matrices are then standardized (**Equation 5**) and multivariate statistical analysis is conducted, resulting in the prediction of putative antimicrobials (classical metabolomics) or putative synergists (interaction metabolomics).

Figure 7. Comparison of the data matrices used for classical metabolomics (A) and interaction metabolomics (B) shown in Figure 6

A. Classical Metabolomics Matrix

Sample	Activity	F1	F2	F3	...	Fm
M1	A_{M1}	$I_{F1,M1}$	$I_{F2,M1}$	$I_{F3,M1}$	⋯	$I_{Fm,M1}$
M2	A_{M2}	$I_{F1,M2}$	$I_{F2,M2}$	$I_{F3,M2}$	⋯	$I_{Fm,M2}$
M3	A_{M3}	$I_{F1,M3}$	$I_{F2,M3}$	$I_{F3,M3}$	⋯	$I_{Fm,M3}$
⋮	⋮	⋮	⋮	⋮		⋮
Mn	A_{Mn}	$I_{F1,Mn}$	$I_{F2,Mn}$	$I_{F3,Mn}$		$I_{Fm,Mn}$

B. Interaction Metabolomics Matrix

Compound Interaction Terms (CIT)

Sample	Activity	F1	F2	F3	...	Fm	F1 × F2	F1 × F3	...
M1	A_{M1}	$I_{F1,M1}$	$I_{F2,M1}$	$I_{F3,M1}$	⋯	$I_{Fm,M1}$	$I_{F1,M1} \times I_{F2,M1}$	$I_{F1,M1} \times I_{F3,M1}$	⋯
M2	A_{M2}	$I_{F1,M2}$	$I_{F2,M2}$	$I_{F3,M2}$	⋯	$I_{Fm,M2}$	$I_{F1,M2} \times I_{F2,M2}$	$I_{F1,M2} \times I_{F3,M2}$	⋯
M3	A_{M3}	$I_{F1,M3}$	$I_{F2,M3}$	$I_{F3,M3}$	⋯	$I_{Fm,M3}$	$I_{F1,M3} \times I_{F2,M3}$	$I_{F1,M3} \times I_{F3,M3}$	⋯
⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮
Mn	A_{Mn}	$I_{F1,Mn}$	$I_{F2,Mn}$	$I_{F3,Mn}$		$I_{Fm,Mn}$	$I_{F1,Mn} \times I_{F2,Mn}$	$I_{F1,Mn} \times I_{F3,Mn}$	⋯

Notes. Each matrix contains data for all mixtures (***M***) in which features (***F***) are detected. A feature (***F***) represents a peak in the LC-MS dataset with a unique m/z value and retention time. The intensity (***I***) of each feature in each mixture is obtained by integrating the relevant selected ion trace in the LC-MS chromatogram. The intensities (***I***) of the features vary between mixtures because the abundance of the compounds (analytes) associated with the features differs between the mixtures. The data matrix used for the interaction workflow (B) includes the same features described for the classical metabolomics matrix (A), but also includes additional compound interaction terms (CIT) (**Equation 3**) for each pair of features detected. The total number of mixtures is ***n*** and the total number of features is ***m***. Each mixture has a measured biological activity (A_{Mn}), which in this study is measured % inhibition against *Staphylococcus aureus*. Compound interaction terms (CIT) are included in matrix B (for interaction metabolomics) and are obtained by multiplying together the intensities (***I***) of the features in a pairwise fashion (**Equation 3**). For classical metabolomics (***A***), the first column on the matrix contains biological activity (A_{Mn}) for each mixture (M_n). For each mixture, there are a series of many additional columns containing the intensity ($I_{Fn,Mn}$) values for each feature (F_n) in each mixture (M_n). Prior to final data analysis, these data matrices are normalized to unit variance as shown by **Equation 5**.

Data Filtering is Important for Interaction Metabolomics

Prior to the data filtering steps used for reducing the number of features (See *Experimental Section*), the number of features detected in the berberine spiked mixtures (M01-M08) was 2894 (**Table 7**). By **Equation 4**, 2894 features would yield 4,189,065 compound interaction terms, which would result in a data matrix with 4,191,959 variables (4,189,065 compound interaction terms plus 2894 features). Data processing with such a dataset would be likely to yield false results because the number of variables is so large compared to the number of biological measurements.[17] To reduce the number of features to a manageable number for an interaction metabolomics workflow, we removed features that were present in the blank and any features for which relative standard deviation of peak area across triplicate analyses was >35%. This reduced the total number of features in the berberine spiked mixtures to 232. We then further filtered the data by removing any features for which peak area did not vary by more than 0.01% across all samples. The rationale for this filtering step the assumption that if biological activity varies across the samples, abundance of the compounds responsible for activity should also vary. The final filtered dataset had 26 features, and the data matrix including features and compound interaction terms had manageable total of 207 variables (**Table 8**). (Note, this number is lower than the 377 variables + features because any feature with peak area of zero will result in a compound interaction term of zero). Similarly, the total number of features in the berberine and piperine spiked mixtures was 33 after all data filtering steps (**Table 8**), and the final compound interaction term data matrix contained a total of 462 variables.

Table 7. Number of features and annotated adducts after the different filtering steps

Data Filtering for M01 to M08			Data Filtering for M09 to M17		
Data Filtering	No. of Features	No. of Annotated Adducts	Data Filtering	No. of Features	No. of Annotated Adducts
Exported from MZmine	2894	38	Exported from MZmine	5698	52
After Blank Filter	446	32	After Blank Filter	1696	51
After RSD Filter	232	30	After RSD Filter	817	50
After variance across samples filter	26	15	After variance across samples filter	33	18

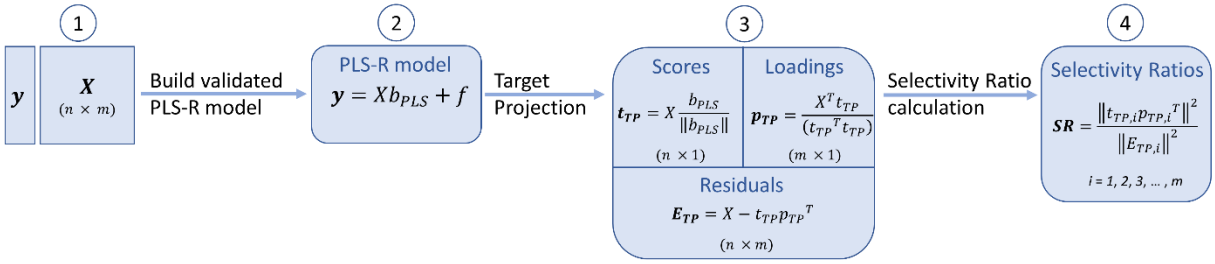
Table 8. PLS Modeling Information

	M09-M17 (Fractions spiked with berberine and piperine)		**M01-M08 (Fractions spiked with berberine)**	
	Classical Model Fig. 5B	**Interaction Model Fig. 5C**	**Classical Model Fig. 5A**	**Interaction Model Fig. 5D**
Object count	9	9	8	8
Variable count	33	462	26	207
Comp. retained	2	3	2	3
Comp. variance (x, y)	1 - 30.50%, 76.47; 2 - 22.27%, 14.26%	1 – 22.33%, 75.74% 2 – 20.44%, 16.25% 3 – 20.38%, 6.48%	1 – 24.42%, 95.89% 2 – 15.00%, 3.18%	1 – 19.61%, 86.37% 2 – 21.79%, 12.27% 3 – 23.06%, 1.03%
Monte Carlo validation threshold	0.357	0.333	0.500	0.500
RMSECV	14.02	11.09	18.61	27.22
R^2Y	0.907	0.985	0.991	0.997
Comp. in SR	2	3	2	3

Comparison of Putative Active Constituents Predicted by Classical Metabolomics and Interaction Metabolomics

For both classical metabolomics and interaction metabolomics (**Figure 6**), PLS was used to determine which features in the chemical dataset were most strongly associated with biological activity, an approach often referred to as "biochemometrics". As with previous biochemometrics studies,[1,2,4,16,17,22,32] we employed the selectivity ratio as a measure of which mixture components associate with biological activity. The selectivity ratio is obtained by dividing the variance explained by the target projection component for a given feature by the residual variance (see *Experimental Section*, **Figure 8**). Because the association between biological activity and ion abundance indicated by selectivity ratios is purely correlative, false correlations may occur. Thus, predictions of biological activity based on selectivity ratio are deemed "putative" and would typically be followed up by a validation experiment testing activity of the isolated compounds. For this study, the biological activity of the mixture components was known a priori so the validity of the predictions from the multivariate statistical model could easily be tested. We predicted that the application of classical metabolomics (**Figure 6A**) to the data for mixtures 01-08 would assign high selectivity ratio to the features associated with berberine, and that application of interaction metabolomics (**Figure 6B**) to mixtures 09-17 would assign high selectivity ratios to the compound interaction terms associated with berberine and piperine.

Figure 8. Calculation of a selectivity ratio



Notes. (1) The data matrix has dimensions of *n* rows and *m* columns. It consists of a response vector ***y*** (biological dataset) with (indicated as $\boldsymbol{A_{Mn}}$ in **Figure 7**) and a corresponding ***X*** predictor matrix. In the case of the conventional workflow, the ***X*** predictor matrix is composed of a series of intensities (***I***) for all detected features (**Figure 7A**). In the case of the synergy workflow, the ***X*** predictor matrix also includes the compound interaction terms (**Figure 7B**). (2) The regression coefficients $\boldsymbol{b_{PLS}}$ from the PLS-R model are used to perform the Target Projection (TP). (3) The TP splits the dataset into predictive loadings $\boldsymbol{p_{TP}}$ and scores $\boldsymbol{t_{TP}}$ and a residual matrix $\boldsymbol{E_{TP}}$ (dimensions in parentheses). (4) For each feature, a selectivity ratio (SR) is calculated.

The effectiveness of the two workflows (**Figure 6**) for predicting active mixture components is compared in **Figure 9**. The plots in this figure show the magnitude of the selectivity ratio on the y-axis calculated for each explanatory variable (LC-MS feature or a compound interaction term) on the x-axis. Note that the selectivity ratio plot includes features detected across all of the samples included in the analysis, and as such are a composite view that differs from the more common way of viewing LC-MS data for each mixture individually.

There is no absolute cutoff value for what qualifies as a "relevant" selectivity ratio. The magnitude of the selectivity ratios obtained in a given analysis will vary depending on the unique characteristics of the dataset being inspected. It is most useful to think of selectivity ratios as a ranking tool. Compounds associated with features that have high selectivity ratios are associated with biological activity and can be prioritized for isolation and further testing. In this case, since we knew for each dataset which compounds were active, we considered the selectivity ratio analysis to be successful when the features or interaction terms with highest selectivity ratios corresponded to those known active compounds.

When the classical metabolomics workflow (**Figure 6A**) was applied to the mixtures spiked only with berberine (M01-M08), one of the two features with highest selectivity ratio corresponded to the $[M]^+$ ion for berberine (m/z 336.1238, 3.28 minutes) (**Figure 9A**). The second of the two features with highest selectivity ratio (m/z 338.1393, 2.89 minutes) was tentatively assigned to the molecule 7,8-dihydroberberine. Inspection of berberine standard by LC-MS (**Figure S6**) indicated the presence of the 7,8-dihydroberberine feature (at 21-times lower intensity than berberine); thus, it appears that 7,8-dihydroberberine is a contaminant of the "pure" berberine. The predicted activity of 7,8-dihydroberberine highlights one limitation of the biochemometrics approach for predicting active mixture components; any feature that correlates with activity will be predicted to be active, but the prediction is purely correlative and must be verified by testing of the compound in isolation. Previous literature has reported that the antimicrobial activity of 7,8-dihydroberberine is similar to that of berberine[40], suggesting that this compound may contribute to the overall activity of this fraction (not evaluated). Several other features are given non-zero selectivity ratios in this analysis (**Figure 9A**). These are likely false positives, but would be deprioritized for isolation given that their selectivity ratio values are small relative to the features associated with berberine and the putative 7,8-dihydroberberine.

One might expect (as we did at the onset of this study), that PLS analysis of a system where activity due to synergy would generate a poor model because compound interaction would be overlooked. Instead, PLS modeling created what would be considered a "good" fit of the data for the mixtures spiked with piperine and berberine (M09-M17, **Figure 9B**). PLS modeling of interaction metabolomics on the mixtures with synergistic effects had a root-mean-standard error for cross-validation (RMSECV) value of 11.09 with an R^2Y of 0.985 (**Table 8**).

Despite the good fit of the model to the experimental data, we know it to be incorrect. Features that correspond to known inactive components of the fractions were incorrectly assigned larger selectivity ratios than berberine or piperine (**Figure 9B**). This appears to be a case of confounding. Thus, the data in **Figure 9B** demonstrate a crucial limitation of classical PLS modeling of metabolomics data. If the observed biological effect is due to synergy, a model that appears to be of high quality can be obtained even though the association pattern between activity and analytes is wrong due to the missing interaction term.

When the new interaction metabolomics approach (**Figure 6B**) was applied to the datasets containing compound interaction terms, the resulting selectivity ratio plot (**Figure 9C**) showed the correct prediction of active constituents. From the 462 features and interaction terms in the dataset (**Table 8**), only five had high selectivity ratios. All of these high selectivity ratio features correspond to the compound interaction terms for piperine features combined with berberine features. Therefore, the problem of confounding demonstrated in **Figure 9C** is resolved when a dataset that includes interaction terms is utilized. Importantly, multiple features are detected in this selectivity ratio plot (**Figure 9C**) because the single piperine molecule forms five different cluster ions in the source of the mass spectrometer (**Figure S5**). Thus, although five compound interaction terms are found, they are redundant in that each corresponds to the

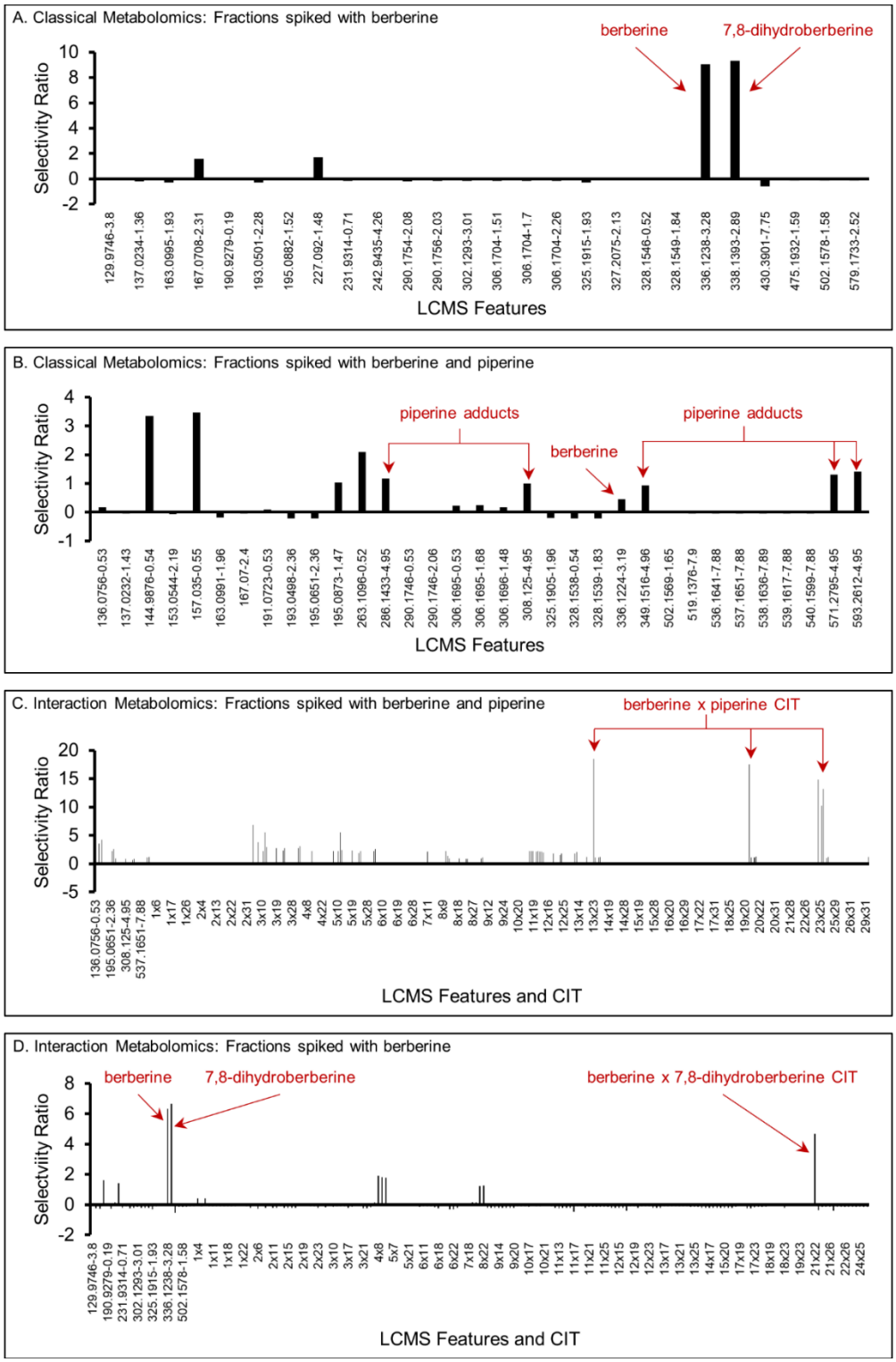
berberine feature area multiplied by the area of a feature representing a different ion formed by LC-MS analysis of piperine.

Because we knew which components in the mixtures were active, it was possible in this study to diagnose the problem of a confounded model (**Figure 9B**) and correct it with inclusion of compound interaction terms (**Figure 9C**). However, an analyst working with a system of unknown composition could waste a great deal of time pursuing putative active compounds with high selectivity ratios in **Figure 9B**, only to discover than none of them are active. Both of the models in **Figure 9B** and **Figure 9C** were of similar quality, having RMSECV of 14.02 and 11.09, respectively, and R^2Y of 0.907 and 0.985, respectively (**Table 8**). Thus, is not possible to diagnose the confounding problem observed in **Figure 9B** based on the model parameters alone.

A possible solution to prevent a false model such as that shown in **Figure 9B** would be to include interaction terms in any metabolomics dataset, preemptively considering possibility that activity is due to compound interactions. To test the effectiveness of such an approach, we applied interaction metabolomics to the dataset for the mixtures containing only berberine (**Figure 9D**), where we knew that activity was not due to compound interactions. A concern prior to carrying out this data analysis was that the inclusion of interaction terms might introduce additional false correlations. However, even with the compound interaction terms, selectivity ratio analysis of the berberine spiked mixtures predicted berberine and putative 7,8-dihydroberberine to be the active constituents (**Figure 9D**), similar to the conclusion for the data matrix without compound interaction terms (**Figure 9A**). These results suggest that interaction metabolomics might be an applicable approach even when compounds in the mixture do not interact.

Notably, and as would be expected, the compound interaction term for the berberine × 7,8-dihydroberberine produced a high selectivity ratio in **Figure 9D**. Thus, the inclusion of compound interaction terms might enable identification of combinations of compounds that interact additively, a concept that would be a good topic of future exploration. Importantly, however, the data in **Figure 9D** demonstrate that a compound interaction term with high selectivity ratio could be indicative of either synergy, additivity, or a false correlation. These results underscore the importance of additional validation for any predictions made based on statistical comparison of chemical and biological datasets.

Figure 9. Comparison of selectivity ratio plots using classical metabolomics (A and B) and interaction metabolomics (C and D)



Notes. The five compound interaction terms (CIT) in Figure 9C are due to the products of berberine and the five piperine adducts: protonated species $[M+H]^+$, sodiated $[M+Na]^+$ species, proton bound dimer, $[2M+H]^+$, sodium bound dimer $[2M+Na]^+$, and a sodiated acetonitrile cluster $[M+ACN+Na]^+$. Specifically, in Panel C**, 13 × 23 =** piperine $[M+H]^+$ × berberine $[M]^+$, **19 × 23 =** piperine $[M+Na]^+$ × berberine $[M]^+$**23 × 24 =** berberine $[M]^+$ × piperine $[M+ACN+Na]^+$, **23 × 32 =** berberine $[M]^+$ × piperine $[2M+H]^+$, **23 × 33 =** berberine $[M]^+$ × piperine $[2M+Na]^+$. In panel D, **21** × **22** = berberine $[M^+]$ × $[M+H]^+$ of putative 7,8 dihydroberberine. The total number of features in the dataset were 17, 32, 528, and 153 for panels A, B, C, and D, respectively. Prior to multivariate statistical analysis, all metabolomics data were filtered based on the requirement that they demonstrate consistent peak area across all replicate analyses, and the requirement that the feature area vary by > 0.01% across all samples in the mixture (see *Experimental Section*).

CHAPTER IV: CONCLUSIONS

Here we demonstrate the utility of interaction metabolomics, a new approach for predicting constituents that interact to exert a combined biological effect. In our experiments, interaction metabolomics enabled identification of synergists, an outcome that was not achieved with a classical metabolomics workflow. While the experiments presented here focused on synergistic interactions, interaction metabolomics could theoretically also be applied to study antagonism, a possibility that could be explored in future studies.

It was critical in this first exploration of interaction metabolomics to apply an experimental design where synergy was known to occur, and where the identities of the compounds responsible for this synergy were also known. Without this knowledge, a negative result could have been interpreted either as no synergistic interactions or a failure of the model to accurately predict synergy. Nonetheless, the scenario for this study was somewhat artificial. In a realistic natural products discovery workflow, synergists might not be present in the same mixtures, or might not be at relevant concentrations to observe interactions. The development of a strategy to overcome these limitations such that interaction metabolomics can be applied to identify unknown synergists would be a worthy topic of future investigation.

The studies presented here suggest that the inclusion of compound interaction terms to account for possible synergistic interactions might be useful as a general practice for metabolomics data processing. The broader applicability and advantages of such an approach need to be evaluated with further investigations using additional experimental systems. Such explorations should, in theory, be quite achievable; any metabolomics experiment could be modified to account for synergistic interactions by appending interaction terms to the data matrix. However, it is important to consider that the inclusion of interaction terms dramatically

increases the number of features in the dataset. Therefore, successful implementation of this approach will require careful attention to ensure that appropriate statistical methods are used and that features lists are filtered to remove noise and reduce them to a manageable size.

REFERENCES

(1) Britton, E. R.; Kellogg, J. J.; Kvalheim, O. M.; Cech, N. B. Biochemometrics to Identify Synergists and Additives from Botanical Medicines: A Case Study with Hydrastis Canadensis (Goldenseal). *J. Nat. Prod.* **2018**, *81* (3), 484–493. https://doi.org/10.1021/acs.jnatprod.7b00654.

(2) Caesar, L. K.; Nogo, S.; Naphen, C. N.; Cech, N. B. Simplify: A Mass Spectrometry Metabolomics Approach to Identify Additives and Synergists from Complex Mixtures. *Anal. Chem.* **2019**, *91* (17), 11297–11305. https://doi.org/10.1021/acs.analchem.9b02377.

(3) Junio, H. A.; Sy-Cordero, A. A.; Ettefagh, K. A.; Burns, J. T.; Micko, K. T.; Graf, T. N.; Richter, S. J.; Cannon, R. E.; Oberlies, N. H.; Cech, N. B. Synergy-Directed Fractionation of Botanical Medicines: A Case Study with Goldenseal (Hydrastis Canadensis). *J. Nat. Prod.* **2011**, *74* (7), 1621–1629. https://doi.org/10.1021/np200336g.

(4) Kellogg, J. J.; Todd, D. A.; Egan, J. M.; Raja, H. A.; Oberlies, N. H.; Kvalheim, O. M.; Cech, N. B. Biochemometrics for Natural Products Research: Comparison of Data Analysis Approaches and Application to Identification of Bioactive Compounds. *J. Nat. Prod.* **2016**, *79* (2), 376–386. https://doi.org/10.1021/acs.jnatprod.5b01014.

(5) Stermitz, F. R.; Lorenz, P.; Tawara, J. N.; Zenewicz, L. A.; Lewis, K. Synergy in a Medicinal Plant: Antimicrobial Action of Berberine Potentiated by 5′-Methoxyhydnocarpin, a Multidrug Pump Inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (4), 1433–1437. https://doi.org/10.1073/pnas.030540597.

(6) Klayman, D. L.; Lin, A. J.; Acton, N.; Scovill, J. P.; Hoch, J. M.; Milhous, W. K.; Theoharides, A. D.; Dobek, A. S. Isolation of Artemisinin (Qinghaosu) from Artemisia Annua Growing in the United States. *J. Nat. Prod.* **1984**, *47* (4), 715–717. https://doi.org/10.1021/np50034a027.

(7) Wani, M.; Taylor, H.; Wall, M.; Coggon, P.; McPhail, A. Plant Antitumor Agents. VI. Isolation and Structure of Taxol, a Novel Antileukemic and Antitumor Agent from Taxus Brevifolia. *J. Anerican Chem. Soc.* **1971**, *93* (9), 2325–2327.

(8) Spelman, K.; Duke, J.; Bogenschutz-Godwin, M. J. The Synergy Principle at Work with Plants, Pathogens, Insects, Herbivores, and Humans. In *Natural Products from Plants*; Cseke, L., Kirakosyan, A., Kaufman, P., Warber, S., Duke, J., Brielmann, H., Eds.; Taylor & Francis: New York, 2006; pp 475–500.

(9) Dettweiler, M.; Marquez, L.; Bao, M.; Quave, C. L. Quantifying Synergy in the Bioassay-Guided Fractionation of Natural Product Extracts. *PLoS One* **2020**, *15* (8 August), 1–12. https://doi.org/10.1371/journal.pone.0235723.

(10) Caesar, L. K.; Cech, N. B. Synergy and Antagonism in Natural Product Extracts: When 1 + 1 Does Not Equal 2. *Nat. Prod. Rep.* **2019**, *36* (6), 869–888. https://doi.org/10.1039/c9np00011a.

(11) Elfawal, M. A.; Towler, M. J.; Reich, N. G.; Golenbock, D.; Weathers, P. J.; Rich, S. M. Dried Whole Plant Artemisia Annua as an Antimalarial Therapy. *PLoS One* **2012**, *7* (12), 1–7. https://doi.org/10.1371/journal.pone.0052746.

(12) Ettefagh, K. A.; Burns, J. T.; Junio, H. A.; Kaatz, G. W.; Cech, N. B. Goldenseal (Hydrastis Canadensis L.) Extracts Synergistically Enhance the Antibacterial Activity of Berberine via Efflux Pump Inhibition. *Planta Med.* **2011**, *77* (8), 835–840. https://doi.org/10.1055/s-0030-1250606.

(13) Tallarida, R. J. Quantitative Methods for Assessing Drug Synergism. *Genes and Cancer* **2011**, *2* (11), 1003–1008. https://doi.org/10.1177/1947601912440575.

(14) Cortina-Borja, M.; Smith, A. D.; Combarros, O.; Lehmann, D. J. The Synergy Factor: A Statistic to Measure Interactions in Complex Diseases. *BMC Res. Notes* **2009**, *2*, 1–7. https://doi.org/10.1186/1756-0500-2-105.

(15) Foraita, R. A Conditional Synergy Index to Assess Biological Interaction. *Eur. J. Epidemiol.* **2009**, *24* (9), 485–494. https://doi.org/10.1007/s10654-009-9378-z.

(16) Rajalahti, T.; Arneberg, R.; Berven, F. S.; Myhr, K. M.; Ulvik, R. J.; Kvalheim, O. M. Biomarker Discovery in Mass Spectral Profiles by Means of Selectivity Ratio Plot. *Chemom. Intell. Lab. Syst.* **2009**, *95* (1), 35–48. https://doi.org/10.1016/j.chemolab.2008.08.004.

(17) Rajalahti, T.; Kvalheim, O. M. Multivariate Data Analysis in Pharmaceutics: A Tutorial Review. *Int. J. Pharm.* **2011**, *417* (1–2), 280–290. https://doi.org/10.1016/j.ijpharm.2011.02.019.

(18) Kurita, K. L.; Glassey, E.; Linington, R. G. Integration of High-Content Screening and Untargeted Metabolomics for Comprehensive Functional Annotation of Natural Product Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (39), 11999–12004. https://doi.org/10.1073/pnas.1507743112.

(19) Lee, S.; van Santen, J. A.; Farzaneh, N.; Liu, D. Y.; Pye, C. R.; Baumeister, T. U. H.; Wong, W. R.; Linington, R. G. NP Analyst: An Open Online Platform for Compound Activity Mapping. *ACS Cent. Sci.* **2022**, *8* (2), 223–234. https://doi.org/10.1021/acscentsci.1c01108.

(20) Nothias, L. F.; Nothias-Esposito, M.; Da Silva, R.; Wang, M.; Protsyuk, I.; Zhang, Z.; Sarvepalli, A.; Leyssen, P.; Touboul, D.; Costa, J.; Paolini, J.; Alexandrov, T.; Litaudon, M.; Dorrestein, P. C. Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. *J. Nat. Prod.* **2018**, *81* (4), 758–767. https://doi.org/10.1021/acs.jnatprod.7b00737.

(21) Inui, T.; Wang, Y.; Pro, S. M.; Franzblau, S. G.; Pauli, G. F. Unbiased Evaluation of Bioactive Secondary Metabolites in Complex Matrices. *Fitoterapia* **2012**, *83* (7), 1218–1225. https://doi.org/10.1016/j.fitote.2012.06.012.

(22) Caesar, L. K.; Kellogg, J. J.; Kvalheim, O. M.; Cech, N. B. Opportunities and Limitations for Untargeted Mass Spectrometry Metabolomics to Identify Biologically Active Constituents in Complex Natural Product Mixtures. *J. Nat. Prod.* **2019**, *82* (3), 469–484. https://doi.org/10.1021/acs.jnatprod.9b00176.

(23) Cech, N.; Junio, H.; Ackermann, L.; Kavanaugh, J.; Horswill, A. Quorum Quenching and Antimicrobial Activity of Goldenseal (Hydrastis Canadensis) against Methicillin-Resistant Staphylococcus Aureus (MRSA). *Planta Med.* **2012**, *78* (14), 1556–1561. https://doi.org/10.1055/s-0032-1315042.

(24) Khan, I. A.; Mirza, Z. M.; Kumar, A.; Verma, V.; Qazi, G. N. Piperine, a Phytochemical Potentiator of Ciprofloxacin against Staphylococcus Aureus. *Antimicrob. Agents Chemother.* **2006**, *50* (2), 810–812. https://doi.org/10.1128/AAC.50.2.810-812.2006.

(25) Kaatz, G. W.; Seo, S. M.; Ruble, C. A. Mechanisms of Fluoroquinolone Resistance in Staphylococcus Aureus. *J. Infect. Dis.* **1991**, *163* (5), 1080–1086.

(26) Clinical and Laboratory Standards Institute. *M07 Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically*; 2018.

(27) Te Dorsthorst, D. T. A.; Verweij, P. E.; Meletiadis, J.; Bergervoet, M.; Punt, N. C.; Meis, J. F. G. M.; Mouton, J. W. In Vitro Interaction of Flucytosine Combined with Amphotericin B or Fluconazole against Thirty-Five Yeast Isolates Determined by Both the Fractional Inhibitory Concentration Index and the Response Surface Approach. *Antimicrob. Agents Chemother.* **2002**, *46* (9), 2982–2989. https://doi.org/10.1128/AAC.46.9.2982-2989.2002.

(28) Van Vuuren, S.; Viljoen, A. Erratum: Plant-Based Antimicrobial Studies Methods and Approaches to Study the Interaction between Natural Products (Planta Medica (2011) 77 (1168-1182)). *Planta Med.* **2012**, *78* (3), 302. https://doi.org/10.1055/s-0031-1298216.

(29) Kaatz, G. W.; Seo, S. M. Inducible NorA-Mediated Multidrug Resistance in Staphylococcus Aureus. *Antimicrob. Agents Chemother.* **1995**, *39* (12), 2650–2655. https://doi.org/10.1128/AAC.39.12.2650.

(30) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinformatics* **2010**, *11*. https://doi.org/10.1186/1471-2105-11-395.

(31) Chau, F. T.; Chan, H. Y.; Cheung, C. Y.; Xu, C. J.; Liang, Y.; Kvalheim, O. M. Recipe for Uncovering the Bioactive Components in Herbal Medicine. *Anal. Chem.* **2009**, *81* (17), 7217–7225. https://doi.org/10.1021/ac900731z.

(32) Kvalheim, O. M. Variable Importance: Comparison of Selectivity Ratio and Significance Multivariate Correlation for Interpretation of Latent-Variable Regression Models. *J. Chemom.* **2020**, No. October 2019, 1–10. https://doi.org/10.1002/cem.3211.

(33) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct. Relationships* **1993**, *12* (1), 9–20. https://doi.org/10.1002/qsar.19930120103.

(34) Henseler, J.; Sarstedt, M. Goodness-of-Fit Indices for Partial Least Squares Path Modeling. *Comput. Stat.* **2013**, *28* (2), 565–580. https://doi.org/10.1007/s00180-012-0317-1.

(35) Kvalheim, O. M.; Arneberg, R.; Grung, B.; Rajalahti, T. Determination of Optimum Number of Components in Partial Least Squares Regression from Distributions of the Root-Mean-Squared Error Obtained by Monte Carlo Resampling. *J. Chemom.* **2018**, *32* (4), 1–12. https://doi.org/10.1002/cem.2993.

(36) Kvalheim, O. M.; Arneberg, R.; Bleie, O.; Rajalahti, T.; Smilde, A. K.; Westerhuis, J. A. Variable Importance in Latent Variable Regression Models. *J. Chemom.* **2014**, *28* (8), 615–622. https://doi.org/10.1002/cem.2626.

(37) Box, G.; Hunter, W.; Hunter, J. S. Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building; John Wiley and Sons: New York, 1978; pp 374–434.

(38) Ljunggren, J.; Bylund, D.; Jonsson, B. G.; Edman, M.; Hedenström, E. Antifungal Efficiency of Individual Compounds and Evaluation of Non-Linear Effects by Recombining Fractionated Turpentine. *Microchem. J.* **2020**, *153*, 104325. https://doi.org/10.1016/j.microc.2019.104325.

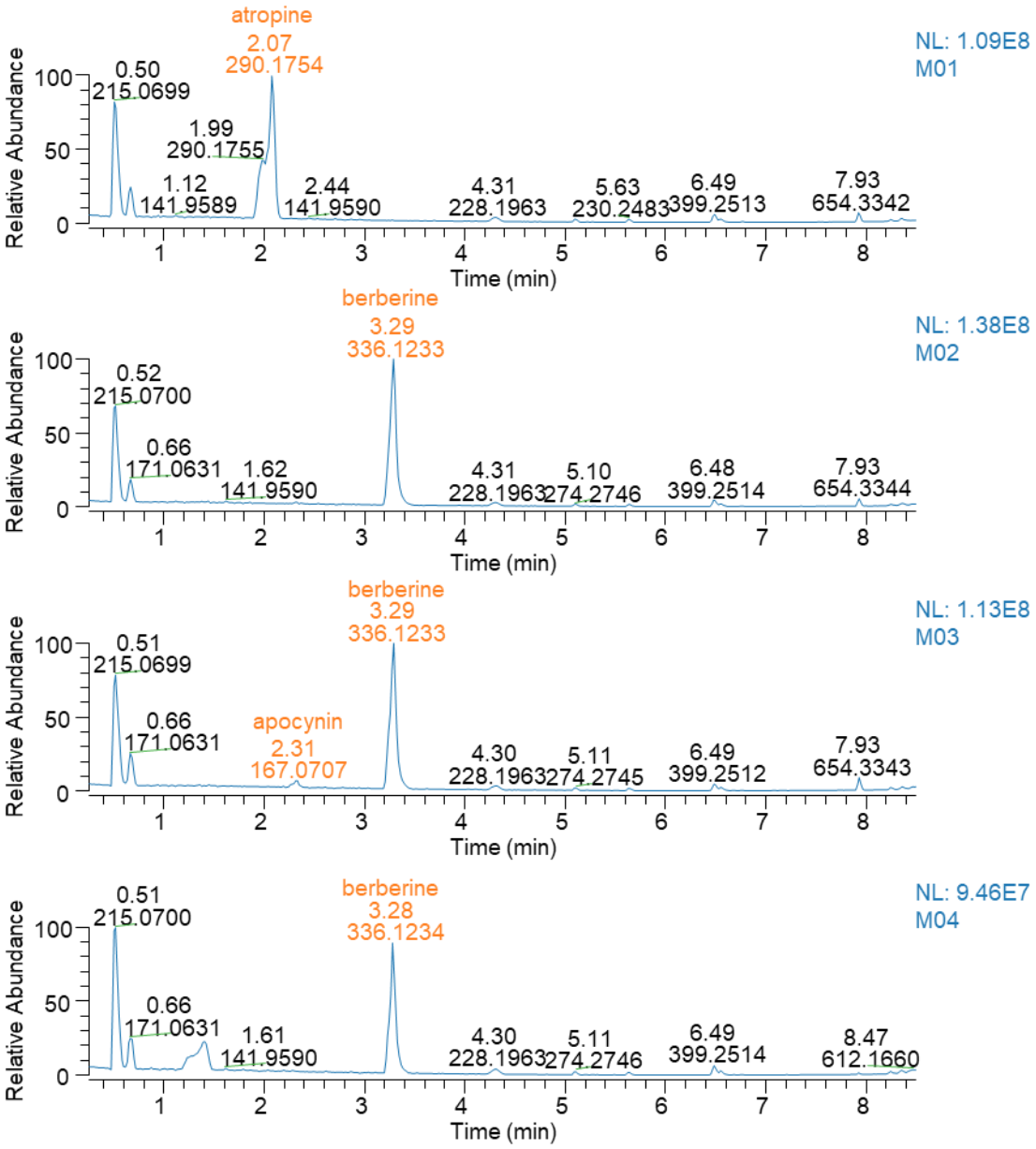
(39) Kalia, N. P.; Mahajan, P.; Mehra, R.; Nargotra, A.; Sharma, J. P.; Koul, S.; Khan, I. A. Capsaicin, a Novel Inhibitor of the NorA Efflux Pump, Reduces the Intracellular Invasion of Staphylococcus Aureus. *J. Antimicrob. Chemother.* **2012**, *67* (10), 2401–2408. https://doi.org/10.1093/jac/dks232.

(40) Olleik, H.; Yacoub, T.; Hoffer, L.; Gnansounou, S. M.; Benhaiem-henry, K.; Nicoletti, C.; Mekhalfi, M.; Pique, V.; Perrier, J.; Hijazi, A.; Baydoun, E.; Raymond, J.; Piccerelle, P.; Maresca, M.; Robin, M. Synthesis and Evaluation of the Antibacterial Activities of 13-substituted Berberine Derivatives. *Antibiotics* **2020**, *9* (7), 1–31. https://doi.org/10.3390/antibiotics9070381.

Table S1. MZmine parameters used for peak picking analysis of the MS raw data

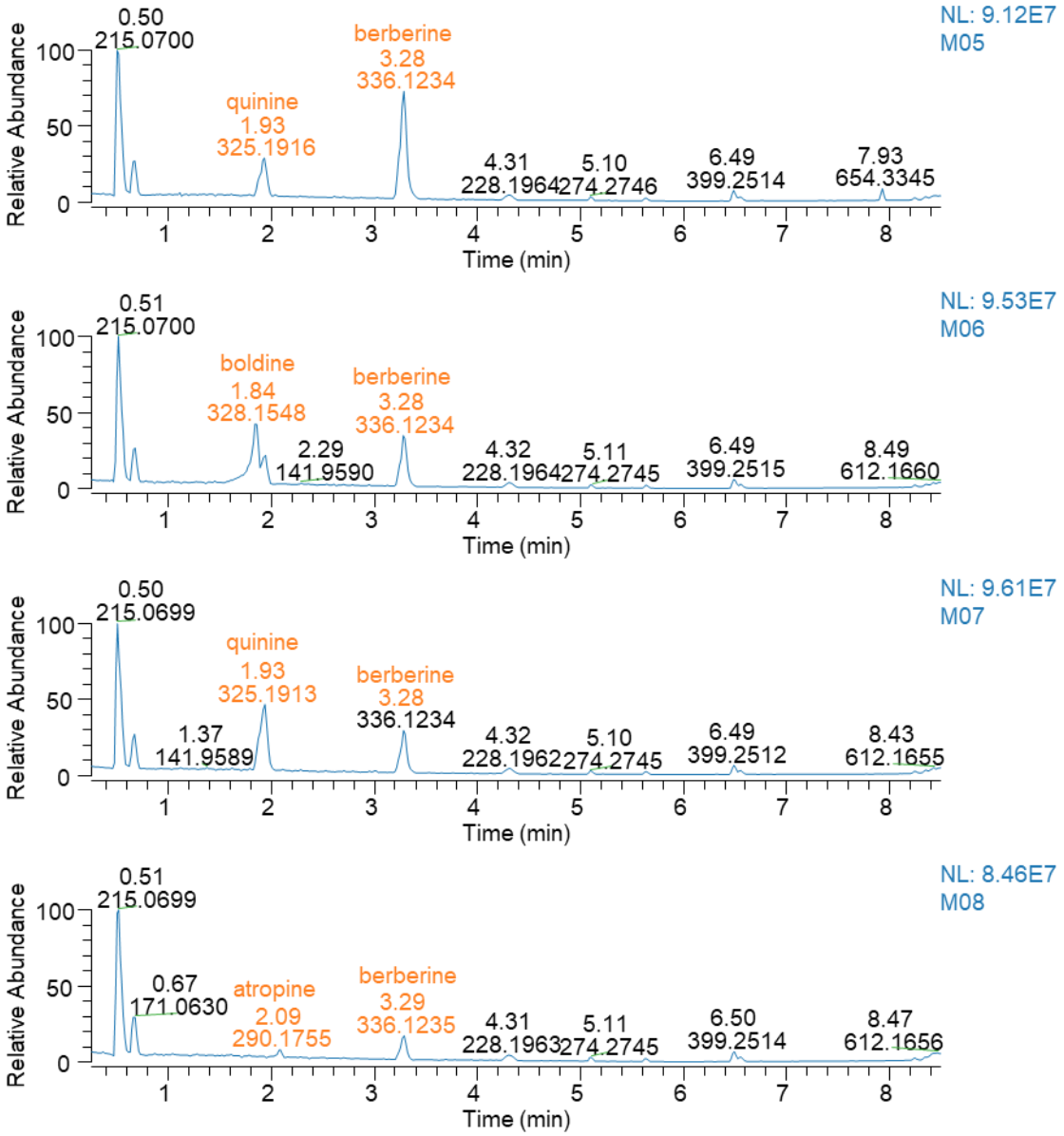
LCMS File No.: 210723 and 220202			By: Warren Vidar
Parameters	**Recommendations**	**210723 (M09-M17)**	**220202 (M01-M08)**
Raw data methods > Filtering > Crop Filter (Optional)			
RT	0.00 to wash time	0.00 to 8.00	0.00 to 8.00
MS level	1	1	1
Polarity	POS / NEG	POS & NEG	POS & NEG
Spectrum type	Profile / Centroid	Profile	Profile
m/z	Auto range	Auto range	Auto range
Raw data methods > Feature detection > Mass Detection			
Mass detector	Exact mass / Centroid	Exact mass	Exact mass
Noise level	5E3 to 5E4	5E3	5E3
Raw data methods > Feature detection > ADAP Chromatogram Builder			
No. of scans	4 or 5	4	5
Group intensity threshold	5E3	5E3	5E3
Min highest intensity	1E5 to 5E5	1E5	1E5
***m/z* tolerance**	0.003 Da	0.003 Da	0.003 Da
Feature list methods > Feature detection > Chromatogram Deconvolution			
Algorithm	Wavelets ADAP	Wavelets ADAP	Wavelets ADAP
***m/z* center calculation**	Median	Median	Median
S/N threshold	10	10	10
S/N estimator	Intensity window SN	Intensity window SN	Intensity window SN
Min feature height	1E5	1E5	1E5
Coefficient/area threshold	30 to 200	50	40
Peak duration range	0.00 to 2.00	0.00 to 2.00	0.00 to 2.00
RT wavelet range	0.00 to 0.10	0.00 to 0.10	0.00 to 0.10
Isotopes > Isotopic Peaks Grouper			
***m/z* tolerance**	0.0015 Da	0.0015 Da	0.0015 Da
RT tolerance	0.05 min.	0.05 min.	0.05 min.
Maximum charge	3	3	3
Representative isotope	Most intense	Most intense	Most intense
Feature list methods > Alignment > Join Aligner			
***m/z* tolerance**	0.0015 Da	0.0015 Da	0.0015 Da
Weight for m/z	2	2	2
RT tolerance	0.05 min.	0.05 min.	0.05 min.
Weight for RT	1	1	1
Require same charge state	checked	checked	checked
Compare isotope pattern	checked	checked	checked
Isotope m/z tolerance	0.0015 Da	0.0015 Da	0.0015 Da
Min. absolute intensity	1E5 to 5E5	1E5 to 5E5	1E5 to 5E5
Isotope pattern min. score	50%	50%	50%
Feature list methods > Gap filling > Same RT and m/z range gap filler			
***m/z* tolerance**	0.0015 Da	0.0015 Da	Peak Finder, 20%
Feature list methods > Filtering > Duplicate filter			
Filter mode	New average	New average	New average
***m/z* tolerance**	0.0015 Da	0.0015 Da	0.0015 Da
RT tolerance	0.05 min.	0.05 min.	0.05 min.
Feature list methods > Filtering > Peak filter			
Height	1E5 to 1E10	1E5 to 1E10	1E5 to 1E10
No. of data points	5 to 100	5 to 100	5 to 100
Identification > Custom database search			
***m/z* tolerance**	0.0015 Da	0.0015 Da	0.0015 Da
RT tolerance	0.02 min.	0.1 min.	0.1 min.

Figure S1. Positive mode full scan base peak chromatogram of M01 to M04



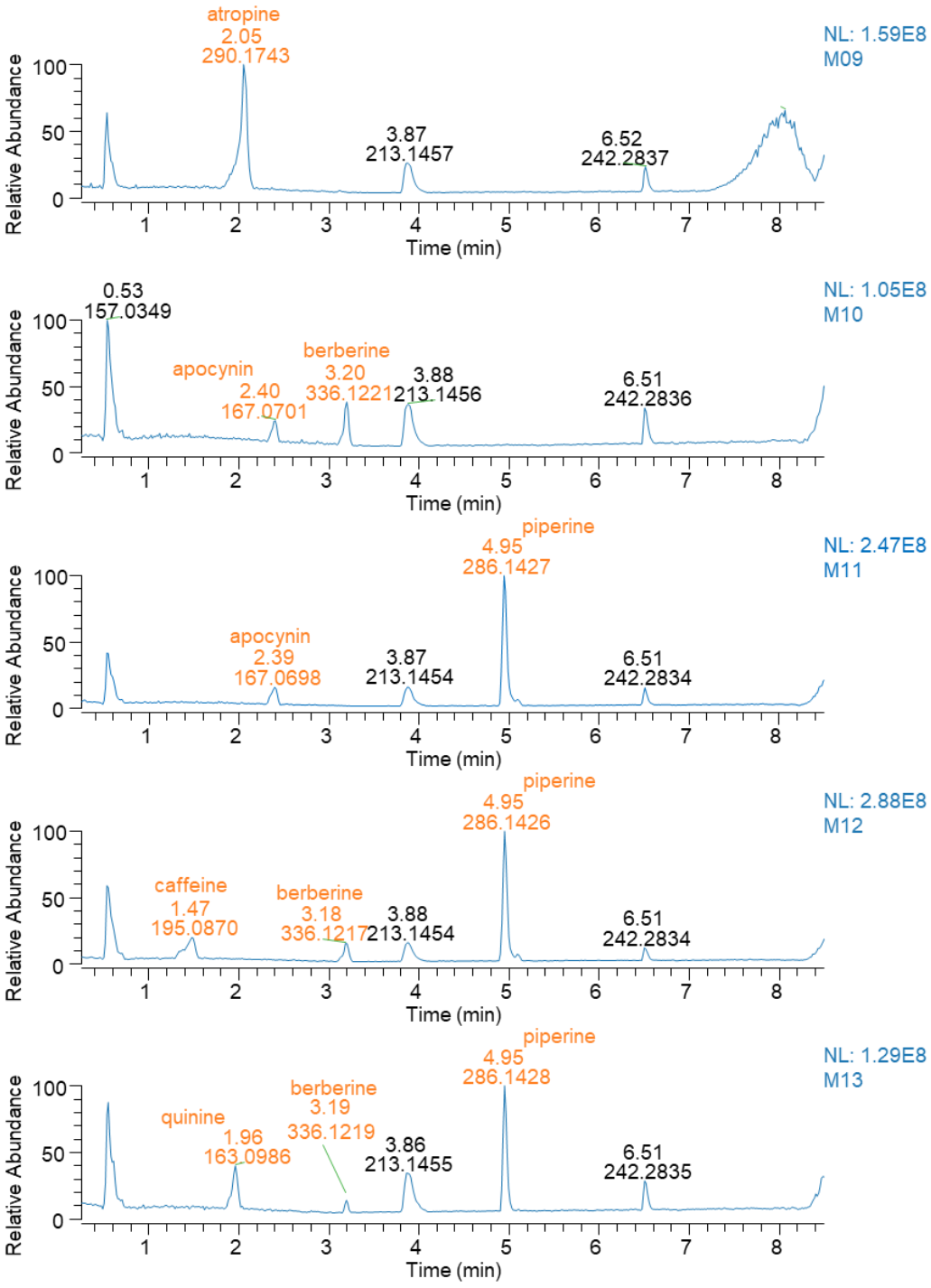
Note. Not all analytes of interest are evident in the base peak chromatogram, but many could be identified with selected ion chromatograms, as indicated in **Table 5** and **Table 6**.

Figure S2. Positive mode full scan base peak chromatogram of M05 to M08



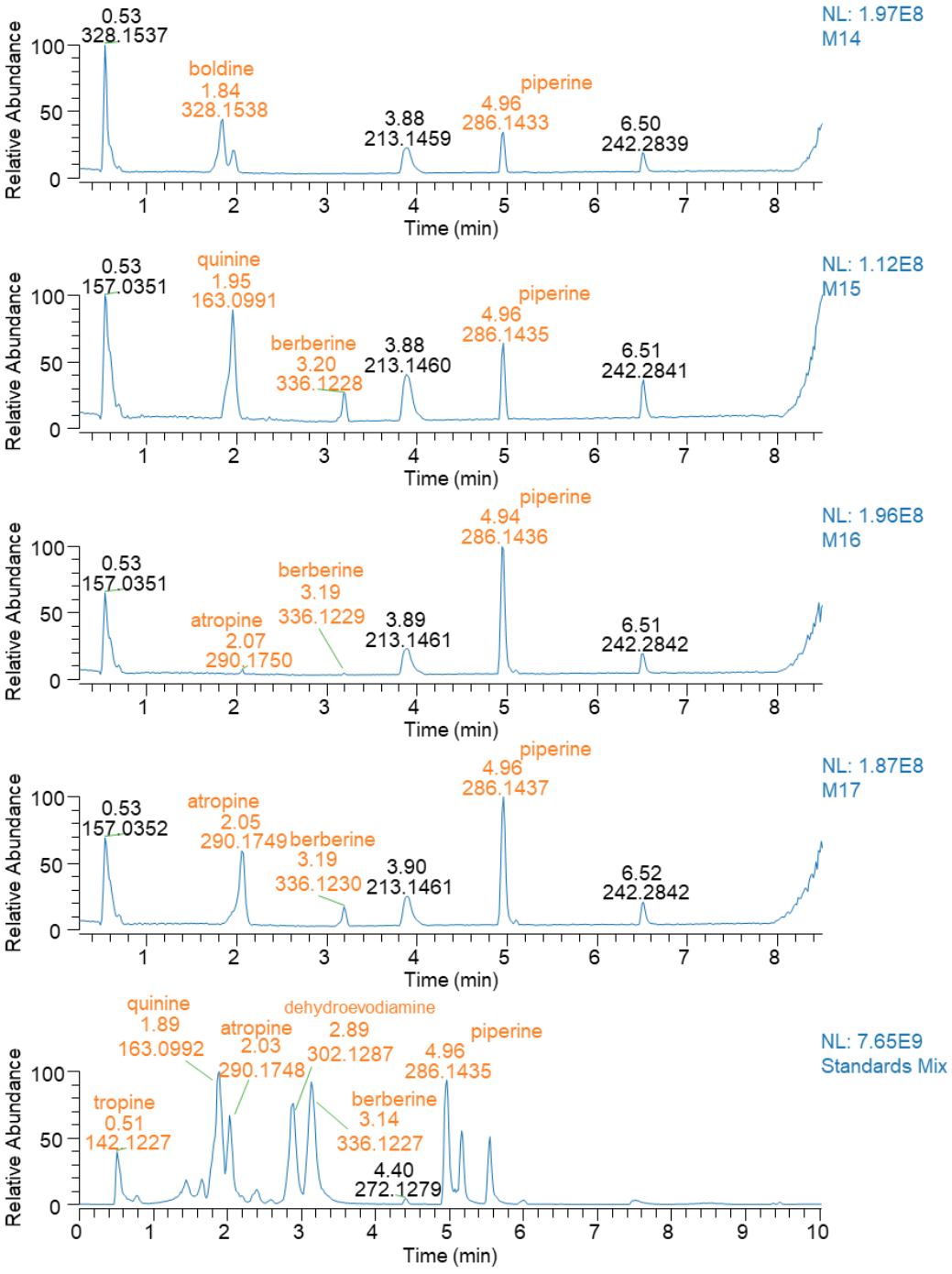
Note. Not all analytes of interest are evident in the base peak chromatogram, but many could be identified with selected ion chromatograms, as indicated in **Table 5** and **Table 6**.

Figure S3. Positive mode full scan base peak chromatogram of M09 to M13



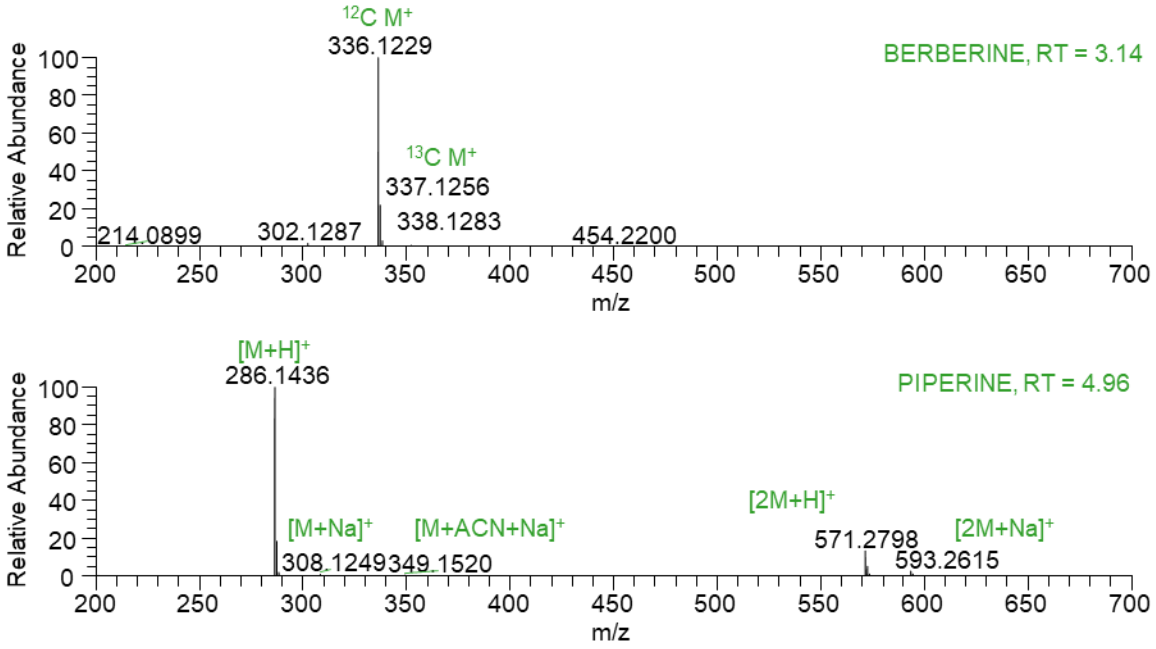
Note. Not all analytes of interest are evident in the base peak chromatogram, but many could be identified with selected ion chromatograms, as indicated in **Table 5** and **Table 6**.

Figure S4. Positive mode full scan base peak chromatogram of M14 to M17 and a mixture of reference standards of compounds used in the simulated extract



Note. Not all analytes of interest are evident in the base peak chromatogram, but many could be identified with selected ion chromatograms, as indicated in Table 5 and Table 6.

Figure S5. Mass spectra of berberine (A) and piperine (B)



Notes. **Figure S5A** shows the ^{13}C isotope of berberine, while **Figure S5B** highlights multiple peaks corresponding to different adducts of piperine. RT refers to the retention time of the peak (time eluted from the LC column) in minutes.

Table S2. List of feature annotations in the LC-MS data

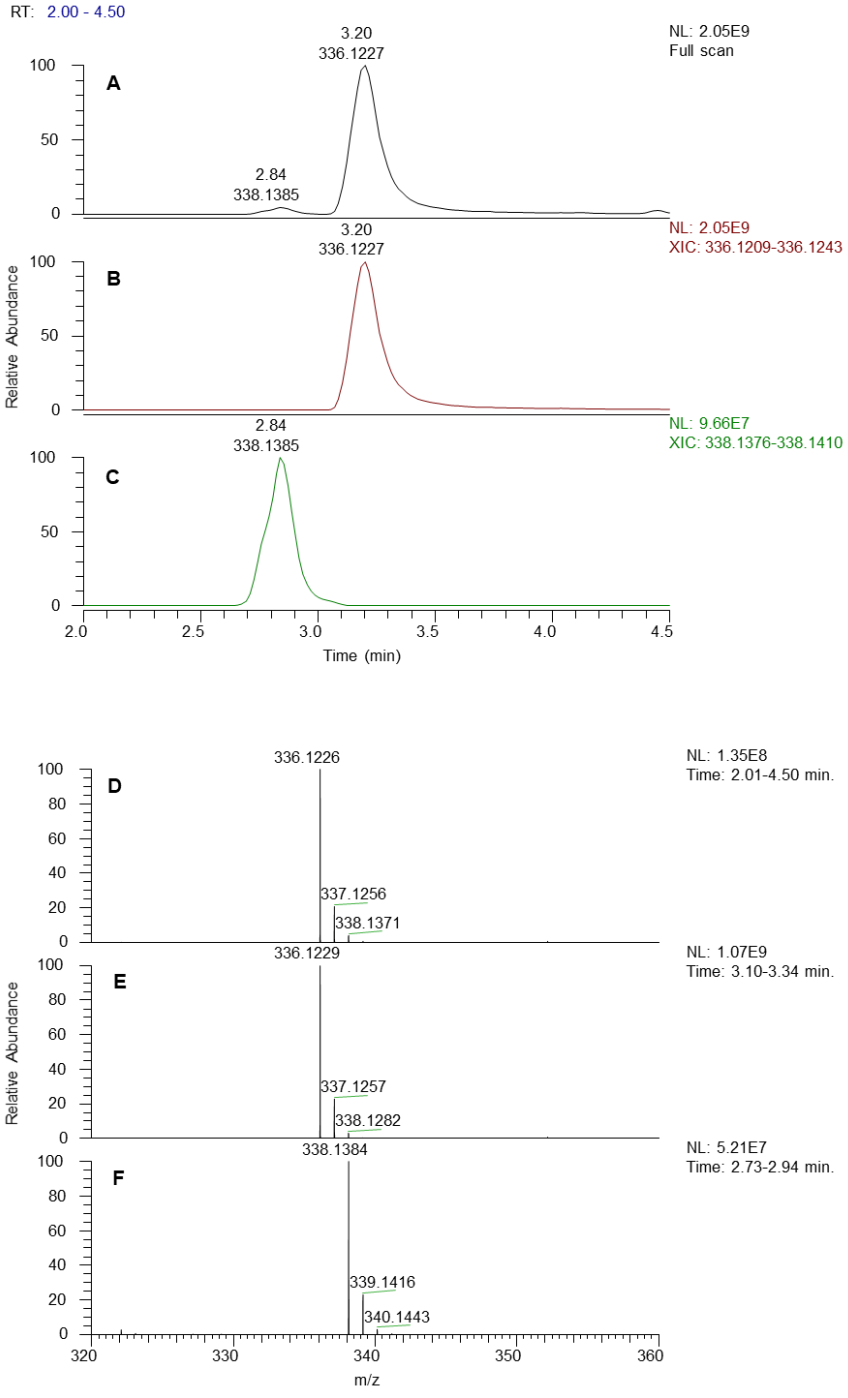
Features (*m/z* -TR)[a]	Molecular Ion or Adduct[b]	Compound Name (Analyte)
137.0232-1.43	$M-H]^-$	3,4-dihydroxybenzaldehyde
139.0389-1.44	$[M+H]^+$	3,4-dihydroxybenzaldehyde
142.1226-0.52	$[M+H]^+$	tropine
151.0388-2.2	$[M-H]^-$	vanillin
153.0544-2.19	$[M+H]^+$	vanillin
154.0862-0.53	$[M+H]^+$	p-octopamine
163.0991-1.96	$[M+2H]^{2+}$	quinine
165.0545-2.4	$[M-H]^-$	apocynin
167.0339-1.73	$[M-H]^-$	vanillic acid
167.07-2.4	$[M+H]^+$	apocynin
169.0497-1.72	$[M+H]^+$	vanillic acid
173-1.41	$[M+Cl]^-$	3,4-dihydroxybenzaldehyde
176.0681-0.51	$[M+Na]^+$	p-octopamine
181.0718-0.78	$[M+H]^+$	theobromine
183.6123-1.94	$[M+ACN+2H]^{2+}$	quinine
193.0498-2.36	$[M-H]^-$	ferulic acid
194.0807-2.2	$[M+ACN+H]^+$	vanillin
195.0651-2.36	$[M+H]^+$	ferulic acid
195.0873-1.48	$[M+H]^+$	caffeine
197.0447-1.79	$[M-H]^-$	syringic acid
199.0599-1.79	$[M+H]^+$	syringic acid
203.0537-0.78	$[M+Na]^+$	theobromine
222.0983-0.78	$[M+ACN+H]^+$	theobromine
229.0268-2.35	$[M+Cl]^-$	ferulic acid
236.0913-2.37	$[M+ACN+H]^+$	ferulic acid
244.0801-0.79	$[M+ACN+Na]^+$	theobromine
251.017-1.44	$[M+TFA-H]^-$	3,4-dihydroxybenzaldehyde
275.0562-1.44	$[2M-H]^-$	3,4-dihydroxybenzaldehyde
286.1433-4.95	$[M+H]^+$	piperine
290.1746-2.06	$[M+H]^+$	atropine
302.1285-2.94	$[M+H]^+$	dehydroevodiamine
304.1919-5.16	$[M-H]^-$	capsaicin
306.1696-1.48	$[M+H]^+$	anisodamine
306.2062-5.16	$[M+H]^+$	capsaicin
308.125-4.95	$[M+Na]^+$	piperine
312.1558-2.06	$[M+Na]^+$	atropine
324.0991-4.95	$[M+K]^+$	piperine
325.1905-1.96	$[M+H]^+$	quinine

328.1539-1.83	$[M+H]^+$	boldine
328.1878-5.16	$[M+Na]^+$	capsaicin
334.1811-0.54	$[2M+3H_2O+2H]^{2+}$	p-octopamine
336.1224-3.19	$[M+H]^+$	berberine
340.1319-1.47	$[M+Cl]^-$	anisodamine
340.1686-5.16	$[M+Cl]^-$	capsaicin
344.1619-5.16	$[M+K]^+$	capsaicin
349.1516-4.96	$[M+ACN+Na]^+$	piperine
350.1612-1.47	$[M+FA-H]^-$	anisodamine
350.1975-5.16	$[M+FA-H]^-$	capsaicin
353.0879-1.36	$[M-H]^-$	chlorogenic acid
355.1019-1.37	$[M+H]^+$	chlorogenic acid
359.1532-1.94	$[M+Cl]^-$	quinine
361.1361-0.78	$[2M+H]^+$	theobromine
369.182-1.93	$[M+FA-H]^-$	quinine
372.1286-1.36	$[M+NH_4]^+$	chlorogenic acid
375.0689-1.36	$[M+Na-2H]^-$	chlorogenic acid
377.0837-1.37	$[M+Na]^+$	chlorogenic acid
383.1182-0.78	$[2M+Na]^+$	theobromine
389.0646-1.36	$[M+Cl]^-$	chlorogenic acid
393.0577-1.35	$[M+K]^+$	chlorogenic acid
399.0931-1.36	$[M+FA-H]^-$	chlorogenic acid
455.3528-7.5	$[M-H]^-$	betulinic acid
456.1505-1.65	$[M-H]^-$	amygdalin
458.1649-1.65	$[M+H]^+$	amygdalin
475.1919-1.66	$[M+NH_4]^+$	amygdalin
480.1472-1.65	$[M+Na]^+$	amygdalin
492.1282-1.66	$[M+Cl]^-$	amygdalin
496.1207-1.65	$[M+K]^+$	amygdalin
498.394-7.5	$[M+ACN+H]^+$	betulinic acid
501.3585-7.5	$[M+FA-H]^-$	betulinic acid
502.1292-1.65	$[M+2Na-H]^+$	amygdalin
502.1569-1.65	$[M+FA-H]^-$	amygdalin
570.1443-1.65	$[M+TFA-H]^-$	amygdalin
571.2795-4.95	$[2M+H]^+$	piperine
579.1724-2.59	$[M-H]^-$	naringin
579.3422-2.06	$[2M+H]^+$	atropine
581.1863-2.59	$[M+H]^+$	naringin
593.2612-4.95	$[2M+Na]^+$	piperine
598.2128-2.59	$[M+NH_4]^+$	naringin
603.1681-2.58	$[M+Na]^+$	naringin
603.2496-2.93	$[2M+H]^+$	dehydroevodiamine
609.2343-4.95	$[2M+K]^+$	piperine

611.4051-5.16	$[2M+H]^+$	capsaicin
615.1492-2.59	$[M+Cl]^-$	naringin
625.1781-2.58	$[M+FA-H]^-$	naringin
633.387-5.15	$[2M+Na]^+$	capsaicin
649.3612-5.16	$[2M+K]^+$	capsaicin
649.3739-1.92	$[2M+H]^+$	quinine
693.1652-2.59	$[M+TFA-H]^-$	naringin
707.1834-1.36	$[2M-H]^-$	chlorogenic acid
731.1777-1.36	$[2M+Na]^+$	chlorogenic acid
911.7138-7.5	$[2M-H]^-$	betulinic acid
913.3073-1.66	$[2M-H]^-$	amygdalin
913.3111-1.64	$[2M-H]^-$	amygdalin
937.3046-1.64	$[2M+Na]^+$	amygdalin
937.3082-1.65	$[2M+Na]^+$	amygdalin
959.3157-1.65	$[2M+FA-H]^-$	amygdalin
1159.3512-2.59	$[2M-H]^-$	naringin
1161.364-2.59	$[2M+H]^+$	naringin
1161.3663-2.59	$[2M+H]^+$	naringin
1205.3577-2.58	$2M+FA-H]^-$	naringin

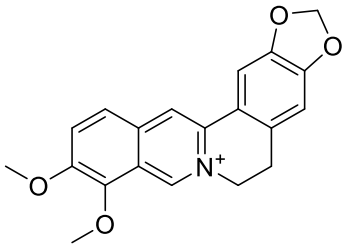
[a] Each feature is described by its measured mass to charge ratio (*m/z*) and retention time (TR, min.)
[b]ACN = acetonitrile, FA = formic acid ,TFA = trifluoroacetic acid,.

Figure S6. Positive mode base peak chromatograms of an isolated berberine standard

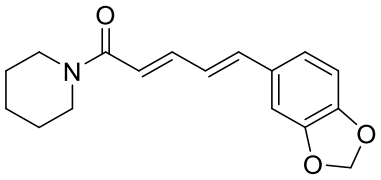


Notes. Figure A shows the full scan within the time range of 2.00 to 4.50 min. Figure B shows the selected ion chromatogram (XIC) for berberine at m/z 336.1227, while Figure C shows the selected ion chromatogram (XIC) for a possible impurity, which is putatively identified as 7,8-dihydroberberine at m/z 338.1385. The two ions have different retention time, making the latter not an adduct or isotope of berberine. Figure D to F show the mass spectra of the full scan and selected ion chromatograms of berberine and putative 7,8-dihydroberberine. NL represents the "normalization level" or intensity used to normalize the abundance on the y-axis.

Figure R1. Chemical structures of berberine (A) and piperine (B).



A



B