

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

BIOACTIVE MOLECULES DETECTION AND IDENTIFICATION IN COMPLEX  
BIOLOGICAL SYSTEMS THROUGH ACTIVITY-BASED METABOLOMICS AND  
PROTEOMICS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

HONGYAN MA  
Norman, Oklahoma  
2020

BIOACTIVE MOLECULES DETECTION AND IDENTIFICATION IN COMPLEX  
BIOLOGICAL SYSTEMS THROUGH ACTIVITY-BASED METABOLOMICS AND  
PROTEOMICS

A DISSERTATION APPROVED FOR THE  
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY THE COMMITTEE CONSISTING OF

Dr. Robert H. Cichewicz, Chair

Dr. George Richter-Addo

Dr. Robert L. White

Dr. Zhibo Yang

Dr. Elizabeth A. Karr



## **Acknowledgements**

Completion of this undertaking could not have been possible without tremendous support and assistance of many individuals. First, I would like to thank my major advisor, Dr. Robert Cichewicz, for his strong support, continuous guidance, and professional advice. I have learned a great deal of knowledge, techniques, and experience under his supervision, which I believe will benefit me in all my future career.

Then I would like to thank my committee members Dr. George Richter-Addo, Dr. Robert White, Dr. Zhibo Yang, and Dr. Elizabeth Karr for their time and intellectual advice to ensure my success in both research and study. I would also like to acknowledge Dr. Steven Foster for his expertise and help in partial LC-MS data collection. Additionally, I want to extend my gratitude to current and former members of the Natural Products Discovery Group. Everyone in this group has provided generous advice and help whenever needed. I could always turn to them when I encountered troubles of all kinds.

I would also like to thank my family for their endless support, love, and especially the warmest encouragement when I was having hard time in the last five years. Last, but not the least for sure, I would like to give most of my appreciation to my loving husband Lin and my cutest and sweetest boys Eric and Evan. They were always there to support and encourage me with kind words. They are the reason I was able to persist in during the most difficult time. I truly believe this would not be done without them.

## Table of Contents

Acknowledgements.....	iv
List of Tables .....	ix
List of Figures.....	x
Abstract.....	xii
Chapter 1: An Introduction to A Critical Analytical Tool: Liquid Chromatography-Mass Spectrometry (LC-MS).....	1
1.1 An Overview of the Principles and Development of Liquid Chromatography-Mass Spectrometry (LC-MS).....	1
1.2 Liquid Chromatography-Mass Spectrometry (LC-MS)-based Metabolomics .....	7
1.3 Liquid Chromatography-Mass Spectrometry (LC-MS)-based Proteomics .....	14
1.4 Perspectives and Future Opportunities .....	20
Acronym list of chapter 1 .....	21
Chapter 2: Chapter Overviews.....	23
2.1 Overview of Graduate Research and Accomplishments .....	23
2.2 Chapter 3. Lickety-Split Ligand-Affinity-Based Molecular Angling System (LLAMAS): A Strategy for Detecting and Dereplicating DNA-Binding Biomolecules from Complex Natural Product Mixtures .....	24
2.3 Chapter 4. Finding Biomass-degrading Enzymes Through an Activity-Correlated Quantitative Protein Profiling Platform (ACPP).....	25
Chapter 3: Lickety-Split Ligand-Affinity-Based Molecular Angling System (LLAMAS): A Strategy for Detecting and Dereplicating DNA-Binding Biomolecules from Complex Natural Product Mixtures .....	26

3.1 Introduction.....	26
3.2 Results and Discussion .....	31
3.2.1 Design and Implementation of an Ultrafiltration-LC-PDA-MS/MS-Based DNA- Binding Assay.....	31
3.2.1.1 Testing Assorted DNA-Binding Agents .....	33
3.2.1.2 Detection of DNA-Binding Molecules in Mixtures and Complex Matrices .....	36
3.2.1.3 Testing LLAMAS 1.0 for the Identification of DNA-Binding Natural Products in a Microbial Extract .....	38
3.2.2 LLAMAS 2.0 for 96-Well Plate-Based Assays.....	41
3.2.2.1 Developing LLAMAS 2.0 .....	41
3.2.2.2 Testing LLAMAS 2.0 Using a Collection of Herbal-Supplement Extracts .....	42
3.2.2.3 Applying LLAMAS 2.0 to Test a Library of Traditional Chinese Medicinal Plant Extracts. ....	44
3.2.3 Conclusions.....	45
3.3 Materials and Methods.....	46
3.3.1 General Experimental Procedures.....	46
3.3.2 Culture and extraction of <i>Streptomyces antibioticus</i> .....	46
3.3.3 Herb extraction.....	47
3.3.4 Lickety-split Ligand Affinity based Molecular Angling System (LLAMAS) .....	47
3.3.5 Untargeted LC-PDA-MS/MS analysis .....	48

3.3.6 Dereplication and characterization of DNA binding candidates .....	49
3.3.7 Fractionation and isolation of DNA binding candidates.....	50
Chapter 4: Finding Biomass-degrading Enzymes Through an Activity-Correlated Quantitative	
Protein Profiling Platform (ACPP).....	52
4.1 Introduction.....	52
4.2 Results and Discussion .....	55
4.2.1 ACPP Development and Proof-of-Concept .....	55
4.2.1.1 HPLC Fractionation.....	56
4.2.1.2 Bioactivity Assay .....	57
4.2.1.3 Proteome Profiling to Generate Protein Elution Profiles.....	59
4.2.1.4 Statistical Correlation Between Activity Pattern and Protein Elution Profiles .....	60
4.2.2 Biomass-degrading Enzymes Characterization from Fungal Secretome using ACPP .....	61
4.2.3 Conclusions.....	65
4.3 Materials and Methods.....	66
4.3.1 A. niger Fermentation and Secretome Extraction.....	66
4.3.2 Anion Exchange LC Fractionation .....	66
4.3.3 Microplate-Based Enzymatic Activity Assays .....	67
4.3.4 LC-MS/MS Analysis .....	68
4.3.5 Peptide and Protein Identification .....	68
4.3.6 Statistical Correlation between Protein Profiles and Activity Pattern .....	69
4.4 Future directions of the ACPP .....	69

Acronym list of chapter 4 .....	71
References.....	72
Appendix.....	82
Supporting Information for Chapter 3 .....	82
List of Contents.....	82
Supporting Information for Chapter 4 .....	112



## List of Tables

**Table 3-1.** Information of 8 known DNA binders..... 34

**Table 4-1.** Manual check of the 4 starch hydrolysis enzyme candidates, which belong to glycosyl hydrolase (GH) family, with correlation R-scores higher than 0.8 in all 156 identified quantifiable proteins in the activity range (e.g. fractions 27-34) ..... 63

## List of Figures

<b>Figure 1-1.</b> The mechanism of electrospray ionization (ESI).....	7
<b>Figure 1-2.</b> Typical workflow of the LC-MS-based metabolomic study.....	8
<b>Figure 1-3.</b> Classical new natural product discovery pipeline.....	11
<b>Figure 1-4.</b> Schematic of the activity-based protein profiling (ABPP).....	20
<b>Figure 3-1.</b> Overview of a strategy designed to detect DNA-binding molecules from chemical mixtures using an ultrafiltration assay coupled to LC-PDA-MS/MS .....	32
<b>Figure 3-2.</b> Methods for the detection of intercalating compounds <b>1-4</b> in the DNA-binding assay .....	35
<b>Figure 3-3.</b> Detection of groove-binding agents <b>5-6</b> and covalent-binding compounds <b>7-8</b> in the DNA-binding assay. ....	35
<b>Figure 3-4.</b> Monitoring the outcome for four DNA intercalators ( <b>1-4</b> ) tested as a mixture in the DNA-binding assay .....	36
<b>Figure 3-5.</b> Detection of DNA binding agents incorporated into a complex soil extract	37
<b>Figure 3-6.</b> Identification of the DNA binding natural products actinomycins D ( <b>9</b> ), V ( <b>10</b> ), and X0 $\beta$ ( <b>11</b> ) from <i>S. antibioticus</i> .....	39
<b>Figure 3-7.</b> Structures of identified DNA binding natural products ( <b>9-18</b> ) from bacterial and plant extracts.....	40

<b>Figure 3-8.</b> Three molecules in the crude extract of root slices of Gold Thread ( <i>Coptis chinensis</i> ) were found to bind with DNA .....	43
<b>Figure 3-9.</b> Confirmation of the DNA binding activities of berberine ( <b>12</b> ), palmatine ( <b>13</b> ), and coptisine ( <b>14</b> ) .....	43
<b>Figure 4-1.</b> Overall workflow of the Activity-Correlated Quantitative Protein Profiling Platform (ACPP).....	56
<b>Figure 4-2.</b> Analysis results of the <i>E.coli</i> cell lysate spiked with standard AAG .....	57
<b>Figure 4-3.</b> Hydrolysis mechanisms of starch and cellulose.....	58
<b>Figure 4-4.</b> The linear dynamic range analysis of the AAG activity using the developed two-step activity assay.....	58
<b>Figure 4-5.</b> Reproducibility performance of the activity assay towards the fractions of the <i>E. coli</i> cell lysate spiked with standard AAG (e.g. 150 µg AAG in 1.85 mg cell lysate).....	59
<b>Figure 4-6.</b> Proof-of-concept results: starch hydrolysis enzyme characterization from the <i>E. coli</i> cell lysate spiked with standard AAG .....	61
<b>Figure 4-7.</b> The results of applying ACPP into biomass-degrading enzymes identification .....	63
<b>Figure 4-8.</b> Cellulose and Cellubiose hydrolysis by commercial β-glucosidase .....	65
<b>Figure 4-9.</b> Co-elution effect evaluation results.....	65

## Abstract

The detection and identification of bioactive molecules (e.g. therapeutic small-molecule lead compounds, functional proteins) from complex mixtures is often a critical step in various biochemical and biomedical research fields. Conventional approaches typically rely on purification of the molecules of interest that is guided by a specific activity/function followed by structure elucidation, chemical/physical property characterization, and bioactivity evaluation. The bioassay-guided purification/characterization process is often conducted in a labor-intensive and low-throughput manner consisting of repeated experimental steps. The lack of tools to directly pinpoint the bioactive molecules from complex crude biological samples is currently a painful challenge in the discovery of new bioactive molecules from biological samples. Emerging LC-MS-based metabolomic and proteomic approaches enabled rapid and high-throughput identification of the proteins and bioactive small molecules from complex biological samples (e.g. natural product extracts/fractions, cell lysates). Nevertheless, it remains challenging to incorporate both the activity and chemical identifications of the bioactive molecules directly from the complex mixtures. To address this problem, we have developed two new approaches: 1) Lickety-split Ligand Affinity based Molecular Angling System (LLAMAS), and 2) Activity-Correlated Quantitative Protein Profiling Platform (ACPP), for direct detection and identification of bioactive small molecules and proteins, respectively, from complex biological samples by combining the activity-based assays with quantitative metabolomics and proteomics studies.

In natural product discovery, bioactive compounds are usually purified via bioassay-guided fractionations. Loss of the activity or failure in purification of the minor or unstable bioactive compounds during the commonly multi-step workflows (e.g. multiple rounds of biological tests and chemical fractionations) is a common and costly limitation in the isolation of bioactive natural

products. A major reason behind the problem lies in that most common bioassay models (e.g. phenotypic and genotypic) are not accessible to the direct detection and identification of individual bioactive chemical component from complex fractions and crude extracts rich in diverse secondary metabolites. To address this, the Lickety-split Ligand Affinity based Molecular Angling System (LLAMAS), an ultrafiltration-LC-PDA-MS/MS-based DNA binding assay, coupled with modern dereplication tools (e.g. GNPS, and DNP), was established for quick and easy identification and dereplication of DNA binding agents in complex samples. This assay was developed using eight known small-molecule DNA binders (**1-8**) with different properties (e.g. DNA binding mechanism, solubility, and LC-MS performance). It was further validated by the successful detection of two DNA binders that were spiked into a highly complex soil extract and the dereplication of three known DNA intercalators including actinomycin D (**9**), V (**10**) and X<sub>0β</sub> (**11**) from the crude extract of *Streptomyces antibioticus*. To improve the throughput, the approach was further optimized to a 96-well plate-based assay—LLAMAS 2.0. Seven DNA binding candidates, including berberine (**12**), palmatine (**13**), coptisine (**14**), fangchinoline (**15**), tetrandrine (**16**), daurisolone (**17**), and dauricine (**18**) were detected and identified from three herbal extracts through high-throughput screening of 394 plant extracts. The results demonstrated that LLAMAS enables efficient identification and dereplication of DNA binding molecules directly from complex metabolite extracts in the early stage of natural product discovery pipelines to avoid repeated purification of known active compounds.

Functional protein (e.g. catalytic enzymes) identification and its activity validation is often a time-consuming and labor-intensive process due to the inclusion of protein overexpression and purification steps. Proteomics approaches have been applied for discovery of novel enzyme candidates from microbial secretomes by comparing protein expression profiles with enzyme activity

of different secretome cocktails obtained under different growth conditions. However, the activity measurement of each enzyme candidate is needed for confident enzymatic activity assignment, which remains to be elucidated. To address this challenge, an Activity-Related Protein Profiling Platform (ACPP) was developed, that can systematically correlates protein-level enzymatic activity patterns with protein elution profiles obtained from the “native” LC prefractionation and quantitative proteomics analysis. Enzymatic activity patterns in sequential fractions measured in different micro-scale bioassays were correlated with protein elution profiles (protein relative intensities in each fractions) using Pearson Correlation algorithm to generate R-scores. The activity was then confidently assigned based on R-score and record of predicted functions in free-accessible databases, such as NCBI and UniProt. The ACPP has been successfully applied to the identification of two types of active biomass-degrading enzymes (e.g. starch hydrolysis enzymes and cellulose hydrolysis enzymes) from *Aspergillus niger* secretome in a multiplexed fashion. By determining protein elution profiles of 156 proteins in *A. niger* secretome, we confidently identified the 1,4- $\alpha$ -glucosidase as the major active starch hydrolysis enzyme ( $R = 0.96$ ) and the endoglucanase as the major active cellulose hydrolysis enzyme ( $R = 0.97$ ). The results demonstrated that the ACPP facilitated the discovery of bioactive enzymes from complex protein samples in a high-throughput, multiplexing, and untargeted fashion.

# **Chapter 1: An Introduction to A Critical Analytical Tool: Liquid Chromatography-Mass Spectrometry (LC-MS)**

## **1.1 An Overview of the Principles and Development of Liquid Chromatography-Mass Spectrometry (LC-MS)**

Liquid chromatography-mass spectrometry (LC-MS) analysis plays significant roles in modern analytical chemistry and bioanalytical chemistry fields. It combines the physical separation techniques of liquid chromatography (e.g. HPLC, UPLC, UHPLC) with mass analysis capabilities of mass spectrometry (MS). As an important and popular analytical tool in laboratories, it has been widely used for both qualitative and quantitative analysis of different biological samples (e.g. metabolites, proteins). With ever-emerging new advanced technologies in column packing and instrumentation, LC-MS has become a prominent and powerful technology with very high specificity, accuracy and sensitivity.<sup>1</sup>

Liquid chromatography (LC) is a useful analytical technique that separates complex mixtures of analytes in solution based on the differential partition coefficient of each analyte between mobile and stationary phase on a chromatography column. Based on different functional chemical groups connected to the stationary phase (e.g. most commonly, silica-based gels), the analytes can be separated by their differential properties, including hydrophobicity [e.g. reversed-phase chromatography (RPC)<sup>2</sup> and hydrophobic interaction chromatography (HIC)], charge [e.g. ion-exchange chromatography (IC)], size [e.g. size-exclusion chromatography (SEC)], affinity (e.g. affinity chromatography), and the combination of hydrophobicity and charge [hydrophilic interaction liquid chromatography (HILIC)]. High-performance liquid chromatography (HPLC) is an advanced type of LC. It typically uses very small packing spherical

silica particles (e.g. 3-10  $\mu\text{m}$  i.d.) as stationary phase with the pressure limit of 5800 psi or 400 bar. For five decades since the 1960s, the pressure limits of HPLC systems remained stagnant, until the "breakthrough" in ultrahigh-pressure liquid chromatography (UHPLC), which holds the pressure up to 1400 bar or 20,000 psi, was reported in 1997, when James Jorgenson packed fused-silica capillaries (30  $\mu\text{m}$  i.d.) with 1.5  $\mu\text{m}$  nonporous octadecylsilane-modified silica particles.<sup>3</sup> A similar study by Milton Lee built on these results in 2001.<sup>4</sup> At that point, the noticeable UHPLC technology still stayed at the "proof-of-concept" stage instead of being applied as a routine analytical tool in laboratories due to the limitation of commercial UHPLC equipment. In 2004, the first commercial Waters Acquity UPLC (Ultra-Performance LC) system was released to the market, together with Acquity UPLC columns (1.0 and 2.1 mm i.d.) packed with sub-2- $\mu\text{m}$  (typically 1.7 or 1.5  $\mu\text{m}$  i.d.) hybrid particles, with an upper pressure limit of 15,000 psi or 1000 bar.<sup>5</sup> Since then, separation science has stepped into a new era, followed by different UHPLC systems emerged by many other manufactures. Compared with HPLC, UHPLC provides faster analysis with good, largely enhanced resolution for the separation of complex samples, and other benefits, such as solvent savings, increased sensitivity, and improved precision for both retention time and peak areas.

Commonly used HPLC and UHPLC detectors include, mass spectrometer (MS), UV detector, charged aerosol detector (CAD), and evaporative light scattering detector (ELSD), but the most popular detectors are UV and MS. In the UV detector, when the sample flows through a detection cell where either single wavelength or full spectra of UV light pass through, the eluent absorbs part of the UV light, which is measured as output for qualitative and quantitative analysis.<sup>6</sup> Photodiode array (PDA) detector is the most popular HPLC detectors which records full



UV-Vis absorption spectra of samples. MS detector measures the mass-to-charge ratio ( $m/z$ ) values of charged particles (ions) in the gas phase.<sup>7</sup> CAD and ELSD are both aerosol detectors. CAD measures charged aerosol particles created via nebulization and positively charged gas collision (e.g.,  $N_2$ ) using an electrometer.<sup>8</sup> ELSD detects the scattering of light caused by the gas-phase analyte particles, which are formed via nebulization and evaporation.<sup>9</sup> Aerosol detectors can detect analytes that are less volatile than the mobile phase, including non- or semi-volatile molecules. They are commonly used to analyze molecules without chromophore, which cannot be detected by UV detectors. In some cases, the LC system is equipped with multiple detectors for specific studies. For example, LC-MS usually combines a UV-Vis detector with a mass spectrometer. Coupling of LC and MS is always desirable and favorable in analytical and bioanalytical fields thanks to the high specificity and sensitivity of MS analysis compared with other chromatographic or aerosol detectors.<sup>10</sup>

Mass spectrometry (MS) is a powerful analytical tool that measures mass-to-charge ratio ( $m/z$ ) values and relative intensities of gas-phase charged ions for the identification and quantification of different biological molecules (e.g. proteins, peptides, lipids, small molecules, etc.). Basically, a mass spectrometer is composed of an ion source, mass analyzer, ion detector, data processor and vacuum system. The Ion source converts analyte particles into ions by bombardment with electrons, molecules, ions, or photons. When passing through the mass analyzer, gas-phase ions with  $m/z$  values are sorted and separated based on the different amounts of deflection that they undergo in the electric and magnetic fields. The most common ion detector used in routine experiments is called electron multiplier. When the beam of ions arrives at the metal plate of the detector, the charge is neutralized by electron transferring, generating electrons, which will be

multiplied and further converted into mass spectra. Signal intensities are related to relative abundance of analyte ions, representing the basic principle of MS-based quantification. Original analyte ions, namely precursor ions or parent ions (MS1) are selected in mass analyzer for fragmentation [e.g. collision-induced dissociation (CID), electron capture dissociation (ECD), and higher-energy collisional dissociation (HCD).] to generate product or daughter ions (MS2) for tandem mass spectra recording. MS2 product ions can be further fragmented to generate MS3, MS4,... MSn spectra. The  $m/z$  values of the product ions (usually MS2 ions) and fragmentation patterns of the analytes are related to unique structure features, providing the foundation of MS-based structure elucidation.<sup>11</sup>

With different mass analyzers, variable resolution and sensitivity can be achieved for the identification and quantification of biological samples. Common mass analyzers include magnetic sector, time-of-flight (TOF), quadrupole mass filter, ion trap, and Fourier transform ion cyclotron resonance (FTICR). The magnetic sector mass analyzer uses a static magnetic field to affect the path of ions in a perpendicular direction to form circular or arched path. Ions with smaller  $m/z$  and higher velocity are deflected more, leading to the separation of ions in the magnetic sector. Herein, the resolution might be limited because ions leaving the ion source may not always have exactly the same kinetic energy and velocity. This phenomenon is very similar to chromatic aberration in optical spectroscopy. To address this, it is usually necessary to add an electric sector to focus the ions, giving them the same starting kinetic energy.<sup>12</sup>

In TOF-MS, ions are first accelerated by an electric field to generate same kinetic energy for those with the same charge.  $m/z$  values of ions are determined by measuring the time taken by ions of different masses to fly from ion source to the detector. Here, the starting time when ions leave ion source needs to be well defined. Thus, in TOF-MS, ions are usually formed by a pulsed

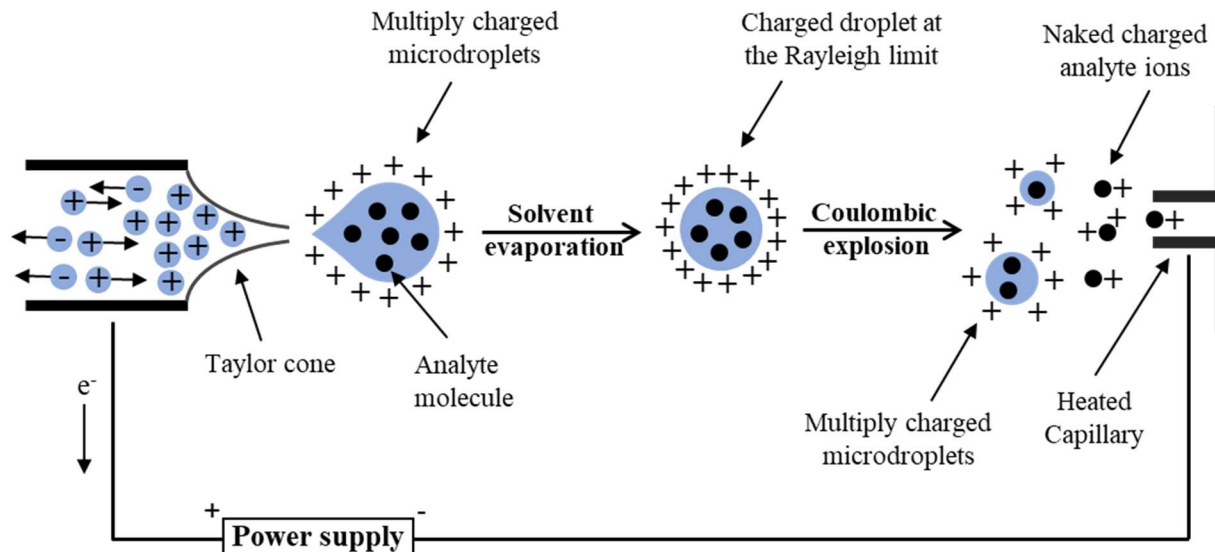
ionization method, such as matrix-assisted laser desorption ionization (MALDI), or rapid switching electric field, releasing the ions from ion source in the manner like a “gate”—open and close alternatively in a very short time. The “Chromatic aberration” phenomenon happens in TOF-MS as well. Ions leaving from ion source don’t always have same starting time or kinetic energy, which affects the resolution. Therefore, to improve the resolution, an ion optic device, so-called “ion mirror” or reflector, is usually adopted to minimize the time or energy difference. Analyte ions with greater kinetic energies will penetrate deeper into the reflector, leading to longer flying time to arrive at the detector.<sup>13</sup>

Quadrupole mass filter selectively allows ions with a specific  $m/z$  value to pass through, by applying combined DC (direct current) and RF (radio frequency) potentials on the quadrupole rods. It is not well suited for pulsed ionization method.<sup>14</sup> Ion trap analyzers can trap ions by bringing them into orbits that are caused by the application of RF fields. They are ideal for MS2 analysis. Selected ions with a specific  $m/z$  value are trapped in orbit, waiting for the introduction of fragmentation. Based on different designs, commonly used ion trap analyzers include linear trap quadrupole (LTQ from Thermo Fisher), orbitrap, cylindrical ion trap, etc..<sup>15</sup>

FTICR-MS measures  $m/z$  values of ions based on their cyclotron frequency in a fixed magnetic field. It provides very high resolution and accuracy, which facilitates the determination of the composition of complex mixtures base on their accurate masses.<sup>16,17</sup> Although different mass analyzers provide different level of resolution and accuracy, we cannot simply judge which one is good for all. Choice of mass analyzer should be based on the application context.

Liquid chromatography-mass spectrometry (LC-MS) is a critical analytical platform that combines advanced separation capabilities of HPLC/UHPLC and specific and sensitive mass detection capabilities of mass spectrometry. Due to its intrinsic detection ability for gas-phase ions,

MS was first coupled with gas chromatography (GC-MS) in the 1950s with commercial instruments available in 1970s.<sup>18</sup> LC-MS development was limited because of the incompatibility of MS ion sources with the continuous liquid stream from the LC end, until the electrospray (ES) ion source was invented by John Bennett Fenn in the 1980s,<sup>19-20</sup> which was awarded Nobel Prize in Chemistry in 2002.<sup>21</sup> In ES ion source, analytes are ionized as shown in Figure 1-1. Specifically, the liquid jet is first nebulized under the effect of electric field and heat, forming a spray of small charged micro droplets. The droplet shrinks due to solvent evaporation, which causes the increasing of charge concentration on the droplets' surface. When the Rayleigh limit is reached, the droplets explosively dissociate because Coulombic repulsion overcomes the surface tension, which is called “Coulombic explosion”, forming smaller, lower charged droplets. The process of shrinking followed by explosion is repeated until individual, charged, 'naked' analyte ions are formed.<sup>22</sup> Electrospray ionization (ESI) is a “soft” ionization technique that is widely used to convert liquid-phase particles from LC end to gas-phase ions, without destroying the intact structure of analytes. It is the most extensively used ion source in laboratories for different biological molecule analysis (e.g. proteins, peptides, small molecules).<sup>23,24</sup> Other ionization methods that can be applied in LC-MS include atmospheric pressure chemical ionization (APCI), atmospheric pressure photoionization (APPI), and fast atom bombardment (FAB). APCI and APPI are complementary to ESI approach for non-polar and thermally stable compounds analysis such as lipids.<sup>10</sup>



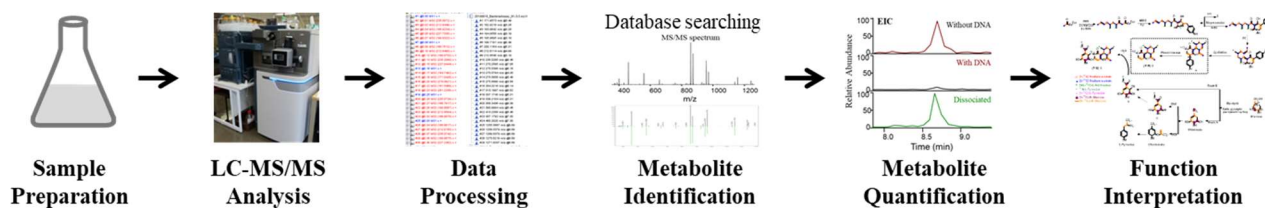
**Figure 1-1.** The mechanism of electrospray ionization (ESI).

Due to its superior capability and robustness in both qualification and quantitation, LC-MS is now successfully applied as a routine analytical and bioanalytical tool in various research areas, such as pharmaceutical,<sup>25</sup> biopharmaceutical,<sup>26</sup> clinical,<sup>10</sup> food, environmental<sup>27</sup> sector, and so on.<sup>28</sup> In this chapter, approaches and applications of LC-MS as a routine metabolomic and proteomic tool will be reviewed.

## 1.2 Liquid Chromatography-Mass Spectrometry (LC-MS)-based Metabolomics

The term “metabolomics” was coined by Nicholson *et al* in 1990s.<sup>29</sup> It represents the comprehensive analysis of a pool of metabolites present intracellularly or extracellularly in a biological system (e.g. cells, biofluids of urine or serum, tissues, organs, organisms). It provides unbiased global qualitative or quantitative assessment of small molecules with low molecular weight (e.g. < 1800 Da).<sup>30,31</sup> A variety of analytical platforms have been applied in metabolomics, including proton nuclear magnetic resonance spectroscopy ( $^1\text{H}$  NMR),<sup>32</sup> MS coupled with

LC, GC,<sup>33,34</sup> capillary electrophoresis (CE),<sup>35,36</sup> and supercritical fluid chromatography (SFC)<sup>37</sup>, among which, NMR, GC-MS and LC-MS are most often used. Each platform has their own advantages and disadvantages, while LC-MS is the most popular technology due to its ease in sample preparation, high sensitivity, and throughput.<sup>38</sup> Metabolomic pipelines can be categorized into targeted and untargeted approaches. The targeted approach is usually hypothesis-driven, focusing on identification or quantitation of a metabolite or a class of metabolites, such as substrates/inhibitors of an enzyme, a group of compounds involved in a specific pathway, and so on. Untargeted analysis measures or profiles all metabolites of a biological system, finding active components (e.g. biomarkers), or generating a new hypothesis for further testing. Typical workflow (Figure 1-2) of the LC-MS-based metabolomic study starts with metabolite extraction and untargeted LC-MS/MS analysis to screen potential or putative compounds of interest, prior to a targeted analysis of metabolite identification and quantitation, followed by function interpretation or pathway analysis.



**Figure 1-2.** Typical workflow of the LC-MS-based metabolomic study

Sample preparation, namely metabolite extraction is usually challenging due to the large chemical diversity, the dynamic nature of chemicals, matrix interference, and sample loss. To extract metabolites from cell or tissue samples, it is required to carefully quench or control the enzyme activities. In some cases, internal standards (e.g. isotope-labeled analytes) are often spiked into the samples during extraction to test if enzyme activity successfully ceases or evaluates the

degradation extent of the target metabolites.<sup>39</sup> Extraction of the target metabolites often needs to be well-tuned and optimized based on their unique properties. Unbiased extraction techniques are often adopted in whole metabolome extraction, including liquid-liquid extraction (LLE), solid phase extraction (SPE), solid-liquid extraction (SLE), protein precipitation, accelerated solvent extraction, etc. Specific protocol optimization should follow the basic principle of compromising minimal matrix interference and maximum sample recovery.<sup>40</sup>

In LC-MS/MS, RPLC and HILIC are commonly used separation techniques for metabolite studies. Selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) is ideal for targeted metabolite investigation, while full scan is required for untargeted analysis. To achieve better chromatographic resolution of co-eluted metabolites and further reduce the sample complexity, additional chromatography techniques are often combined with LC-MS, such as IE and SEC. Recent development of UHPLC also facilitate the improvement of peak capacity and chromatographic resolution.

In data analysis, LC-MS data usually needs to be converted into a peak list for further metabolite identification and quantification, requiring several steps of pre-processing, including peaking filtering, baseline correction, peaking detection and matching, and retention time alignment. Besides commercially available software provided by most instrument manufactures, there are some publicly available tools for LC-MS data pre-processing, including Mzmine,<sup>41</sup> XCMS,<sup>42</sup> MetAlign,<sup>43</sup> and many others.

Metabolomics studies can be divided into qualitative analysis, wherein molecules are identified, and quantitative analysis leading to relative or absolute quantification of metabolites. Metabolite identification is another major challenge in metabolomic studies.<sup>44</sup> Currently, metab-

olite identification is achieved by MS and MS/MS spectra searching against open-source databases [e.g. Human Metabolome Database (HMDB),<sup>45</sup> MassBank,<sup>46</sup> METLIN,<sup>47</sup> and ChemDB<sup>48</sup>] and in-house libraries, followed by manual verification. Large chemical and physical diversities of metabolites (sugar, lipids, steroids, amino acids, etc.) make it very difficult to derive general rules for fragmentation prediction. Due to the huge population of metabolites, and ever-emerging new compound discovery, the lack of standards and MS database is a major bottleneck in metabolite annotation.<sup>38, 49</sup> Sometimes, a compound can be identified as an analogue of a known molecule, yet the exact structure cannot be determined through MS spectra comparison only.

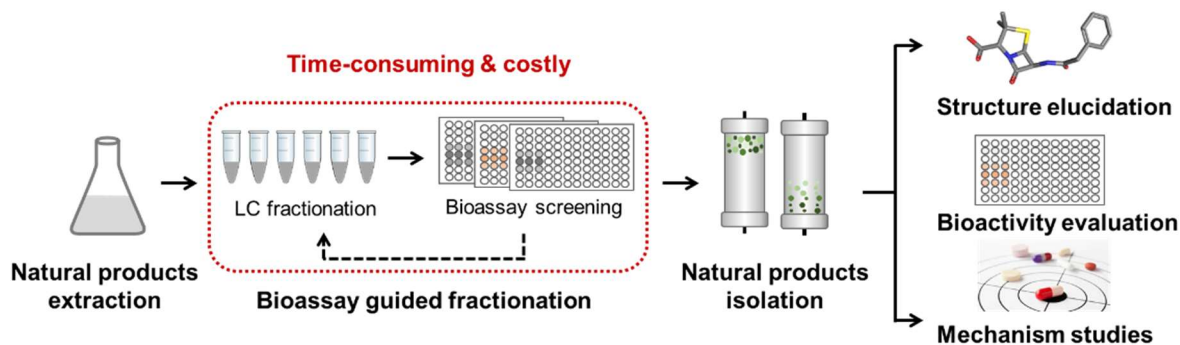
Current quantification strategies in metabolomics are essentially targeted metabolite quantification and untargeted metabolomic profiling. The MRM technique serves as a foundation of high-quality metabolite quantification, while the full scan technique is utilized for metabolome profiling. In some cases, relatively large-scale but not whole metabolome quantitative analysis is needed, wherein MRM is limited by its relatively low metabolite coverage and throughput capacity. With the advancement and improvement in mass spectrometer, new approaches then emerged, bridging this gap. For instance, Asara's group successfully established a platform for large-scale metabolite (258 metabolites) analysis, using a MS platform of triple quadrupole for MRM analysis, with faster scan speed (e.g. 3-ms for dwell time) and the capability of fast positive-negative ion mode switching.<sup>50</sup> In addition, parallel reaction monitoring (PRM) coupled with high-resolution MS (e.g. Orbitrap) is an alternative platform for large-scale metabolite quantification.<sup>51</sup> MRM is powerful due to its dramatically fast scan speed and polarity switching, while PRM benefits from its high resolution and full MS/MS acquisition for higher confidence identification of metabolites.



Untargeted metabolomics often involves vast data processing, which requires fast and accurate bioinformatic tools to deal with the high complexity of the generated metabolomic data. Commonly used statistical tools in metabolomics are t-test, analysis of partial least squares discriminant analysis (PLS-DA), principal component analysis (PCA), analysis of variance (ANOVA), and so on. Multiple web-based untargeted metabolomic platforms have been established, allowing users to update and process their data in a few mouse clicks. Some platforms will be reviewed later in this chapter.

LC-MS-based metabolomics has been demonstrated to be a robust analytical tool in multiple areas, such as drug discovery,<sup>52,53</sup> metabolic pathway analysis,<sup>54,55</sup> and identifying biomarkers of various diseases, including hepatocellular carcinoma (HCC),<sup>56</sup> colorectal cancer,<sup>57</sup> coronary heart disease,<sup>58</sup> insulin resistance,<sup>59</sup> and many others. Here, the application of LC-MS based metabolomics in drug lead discovery from natural sources is reviewed in detail.

Natural products (NPs) are usually refer to secondary metabolites produced by a living organism (e.g. fungi, bacteria, plant), which are considered as an important source for new drug discovery.<sup>60,61</sup> The classical natural product discovery pipeline starts with activity screening of crude metabolite extracts, followed by compound isolation using bioassay-guided fractionation technique, structure elucidation, and activity measurement (Figure 1-3).



**Figure 1-3.** Classical new natural product discovery pipeline

Although successful, discovery of new bioactive metabolites (e.g. drug leads) is always challenging due to the repeated isolation of known compounds. To overcome this problem, metabolomics was introduced to facilitate the compound dereplication in a high throughput manner, by providing a better view of vast data and broader metabolome coverage.<sup>62</sup> Moreover, MS/MS fragmentation provides additional information for structure elucidation. Various powerful untargeted metabolomic platforms have been developed to successfully facilitate the bioactive natural product isolation, including structure dereplication and activity prediction, which are the major applications in natural product drug discovery. These platforms include XC-MS online,<sup>63</sup> MZmine 2,<sup>64</sup> OpenMS,<sup>65</sup> MS-DIAL,<sup>66</sup> MAVEN,<sup>67</sup> W4M,<sup>68</sup> MetaboAnalyst,<sup>69</sup> Molecular Networking,<sup>70,71</sup> and many others. Primarily, they utilize feature (e.g. retention time, MS1, MS2, etc.) detection/alignment algorithms coupling with other statistical approaches to discover known compounds or predict the active components. The dereplication process is generally achieved by direct matching, scoring, and grouping of the characteristic experimental data of a test compound with those deposited in natural product databases (e.g. Dictionary of Natural Products (DNP), MarinLit, Antibase, DEREK-NP, and Atlas.)<sup>72-74</sup> Molecular Networking platform (<http://gnps.ucsd.edu>) implements LC-MS/MS molecular networking as a cheminformatic tool and allows users to annotate MS/MS data with structures in a rapidly expanding public database, making it advantageous for dereplication query. This platform automatically associates MS/MS spectra based on similarity of fragmentation patterns, with the underlying concept that structurally related molecules will fragment in related ways to give analogous patterns. These relationships are used to define molecular networks, in which the fragmentation patterns of related structures are clustered, analyzed, and visualized. Thus, new analogues of a known class of structures can be predicted. Several studies witnessed the successful application of Molecular Networking

in active compounds isolation. The Yang and Dorrestein groups isolated several new antibacterial amino-polyketide derivatives, vitroprocines A-J, from the marine bacterium *Vibrio sp.*, facilitated by molecular networking.<sup>75</sup> Similarly, Gerwick's group isolated anti-inflammatory and analgesic sphongonucleosides from Caribbean sponge *Tectitethya crypta*,<sup>76</sup> and purified new linear lipopeptides, microcolins from marine cyanobacterium *Moorea producens* most recently.<sup>77</sup> Molecular Networking platform was improved recently to accommodate the activity prediction function for drug lead discovery. Bioactivity-based Molecular Networking is capable of finding candidate active compounds from bioactive fractions, through the correlation between the activity patterns and the elution profiles of all metabolites detected in LC-MS analysis.<sup>78</sup> Another metabolomic platform MetaboAnalyst can also differentiate active and inactive molecules via the *mummichog* algorithm, which allows direct prediction of pathway activities of molecules from high-resolution MS peaks, without the prerequisite of accurate peak annotation.<sup>69</sup> Espindola's group successfully annotated active compounds acetogenins in the extract of Brazilian plant *Annona crassiflora* via MetaboAnalyst analysis.<sup>79</sup>

LC-MS based metabolomics has shown its success and importance in various fields, yet there is still room to improve in several areas, like sample preparation, metabolite identification, and activity prediction. Efforts should be made urgently to enlarge the metabolite MS library, develop more advanced bioinformatic tools that can link the metabolomic dataset and the activity results. Intergradation of metabolomics and genomics for better understanding of the roles of the molecule of interest in various pathways might be a feasible and promising direction.

### **1.3 Liquid Chromatography-Mass Spectrometry (LC-MS)-based Proteomics**

Proteins are critical functional components in different cellular activities (e.g. signal transportation, product manufacture, self-protection, and routine maintenances).<sup>80,81</sup> Due to these important functional roles, comprehensive studies of the proteins of interest is paramount to understand the underlying mechanisms of these processes. Conventional approaches for studying functional proteins mainly rely on targeted gene overexpression, followed by the purification of the protein of interest and related structural studies and activity assays.<sup>3,81,80,82</sup> are typically time consuming and low throughput. Proteomics, a large-scale analysis of proteome (e.g. the entire set of proteins that generated by an organism, cell, tissue, or system), relying on mass spectrometry analysis, provides a platform for protein characterization and quantification in complex samples. By coupling with LC separation, proteomics has been processed into a high-throughput and large-scale approach, as modern advanced LC techniques allows thousands of proteins to be analyzed in a single LC/MS run.<sup>83</sup>

The workflow of LC-MS-based proteomics often involves sample preparation, LC-MS/MS analysis, and data analysis. The involved MS instrumentation and LC separation tools are similar as those adopted in metabolomics, so this part will not be reviewed much here. Based on different protein characterization purpose and upstream sample preparation, proteomics can be categorized into bottom-up and top-down strategies. Basically, bottom-up proteomics, namely shotgun proteomics, refers to chemical or enzymatical cleavage of the protein into peptides for further LC separation and MS analysis, while top-down proteomics directly fractionate and detect the intact proteins. Both approaches are applicable for protein identification and quantification, yet top-down proteomics is preferable for the identification of various proteoforms, which

are critical to functions. Proteoforms of proteins often result from combining regulations of genetic variation, alternative splicing, and post-translational modifications (PTMs).

As the most commonly used proteomics platform, bottom-up proteomics is well-established for proteome profiling and quantification.<sup>84</sup> In sample preparation stage, desalting is required for the digested peptide samples to accommodate the interface of LC-MS, because the salts introduced in the digestion step is non-volatile, which will accumulate in the end of the capillary tip, causing clog issues. The most often used enzymes in protein digestion are trypsin and pepsin. Solid-phase extraction (SPE) is a routine desalting technique for peptide sample clean-up prior to LC-MS analysis. Integrated LC-ESI-MS is the most popular tool in high throughput proteomic analysis of complex samples. In LC-MS-based proteomics, upstream separation or fractionation techniques are critical for achieving improved analytical dynamic range and broader proteome coverage. Common separation techniques including one-dimensional (1D), two-dimensional (2D), multi-dimensional LC, and most recently emerged ion mobility spectrometry. They are all widely used for global and quantitative proteomics analysis. RPLC (e.g. C18-based stationary phase) is a routine couple of MS detection due to the compatibility of its mobile phase. To improve the peak capacity, researchers often resort to extending the length of capillary column (e.g. 80 cm, 200 cm) with long separation gradient (e.g. 180 min, 33 hours).<sup>85,86,87</sup> UHPLC, referred to the discussion in 1.2, is another mature separation technique with higher peak capacity, better chromatographic resolution, facilitating the characterization and quantification of low abundance proteins in targeted or global proteomic analysis.<sup>88,89-90</sup> To fulfill the higher requirements of accuracy and sensitivity for highly complex sample analysis, the hybrid platforms of 2D or multi-dimensional separations towards the peptide samples have been developed and applied with success. Considering the compatibility, the separation technique directly coupled with

MS is usually RPLC, while the upper stream fractionations can be strong cation exchange (SCX),<sup>91-93</sup> anion exchange (AEX),<sup>94</sup> HILIC,<sup>95</sup> SEC,<sup>96</sup> capillary isoelectric focusing (CIEF),<sup>96-97</sup> and so on. High-pH-RPLC coupling low-pH-RPLC with fraction concatenation is a non-typical 2D proteomic analysis platform, yet affording better proteome coverage.<sup>98</sup> Ion mobility spectrometry (IMS) is an analytical tool that separates gas-phase ions based on their differential mobility, which is caused by different molecular size and shape, through a buffer gas.<sup>99</sup> It can resolve the ions that are indistinguishable by MS only, such as isomers, or gain more insights into conformational structural information.<sup>99</sup> It has been applied for analyzing small molecules (e.g. < 500 Da),<sup>100</sup> peptides,<sup>101-102</sup> proteins, and protein complex under native conditions<sup>103-105</sup>.

Protein identification is accomplished through comparing the collected MS and MS/MS spectra with the predicted or previously identified features in a library (e.g. UniProt and NCBI), using bioinformatic software, like MS-GF+,<sup>106</sup> Sequest,<sup>107</sup> X!Tandem,<sup>108</sup> Mascot,<sup>109</sup> commercially available software (e.g. Proteomic MS Solutions from Thermo, Progenesis QI from Waters), and many others. Compared with small molecule identification in metabolomics, protein/peptide identification tools are mature and well-established, with handy well-predicted databases, thanks to the advanced genome sequencing technology.

Quantitative proteomics is a critical analytical tool in system biology. The quantitation approaches can be categorized into label-free and stable isotope labeling strategies.<sup>110</sup> Using a label-free strategy, peptide quantification is achieved either by comparing the extracted MS1 peak intensities/areas, or the total acquired MS2 spectra number of the same peptide across different samples, which is often referred as spectral counting approach. Various open-source and commercial data processing software have been developed for label-free quantitative proteomics analysis, including MZmine, MSight, MsInspect, MapQuant, OpenMs, PEPPer, SuperHirn,

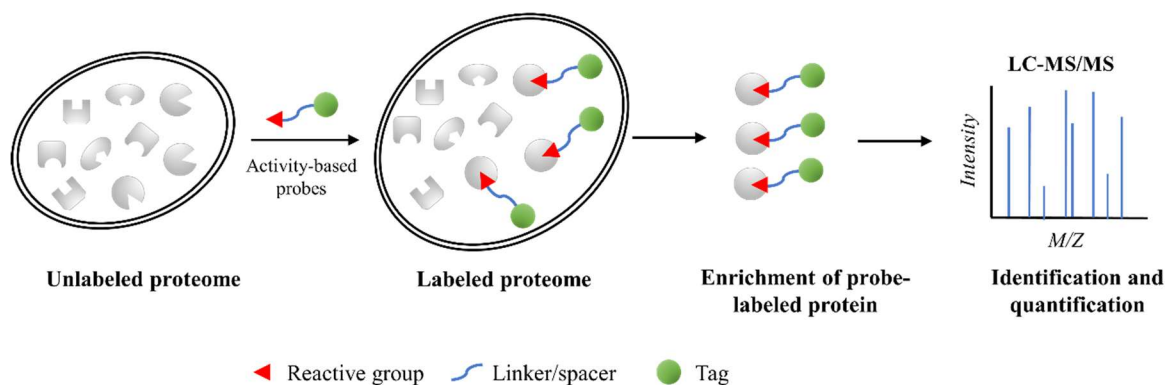
Skyline, and many others.<sup>111-113</sup> Various stable isotope labeling approaches are available for bottom-up quantitative proteomic studies, including Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC), Tandem Mass Tag (TMT), Isobaric Tags for Relative and Absolute Quantification (iTRAQ), Isotope-Coded Affinity Tag (ICAT), Isotope Coded Protein Labeling (ICPL), <sup>15</sup>N/<sup>14</sup>N metabolic labeling, <sup>18</sup>O/<sup>16</sup>O enzymatic labeling, and many other chemical labeling approaches. Both label-free and isotopic labeling strategies have their own merits and demerits. Label-free approaches are simple and robust yet limited by the linear dynamic range and accuracy for low abundance protein quantification. The isotopic labeling strategies may address these, yet still have several potential limitations, including complex sample preparation and handling, expensive labeling reagents, incomplete labeling issues, and the requirement of specific quantification software.<sup>114</sup>

In last several decades, bottom-up proteomics has been extensively utilized in system biology. Nevertheless, it is limited by PTMs identification (e.g. phosphorylation, glycosylation, acetylation, and methylation) as their position information is lost in the protein chopping procedure during enzymatic digestion. Top-down proteomics seeks to eliminate these problems by detecting intact protein samples. This approach allows for 100% sequence coverage and full characterization of proteoforms.<sup>115</sup> Despite this advantage, top-down proteomics is more challenging due to wide range of the protein sizes and various technical difficulties, including sample handling, MS instrumentation for large *m/z* values resolving, and downstream data analysis.<sup>116</sup> Upstream separation techniques could be RPLC (e.g. C3-5 stationary phase), SEC, EX, HILIC, HIC, CIEF, which are similar to those in bottom-up proteomics analysis. On the MS side, the detection and identification of intact proteins relies on high resolution mass spectrometers (e.g. TOF, Orbitrap, and FTICR), especially for proteome-wide level studies.<sup>117</sup> The commonly used

data analysis software are ProSight, TopPIC, Big Mascot, PIITA, and many others. Compared with quantitative bottom-up proteomics, high throughput large-scale proteoform profiling and quantification has been left behind, yet due to the ever-emerging progress and improvement in LC separation techniques, MS instrumentation, ionization method [e.g. UV photodissociation, (UVPD)],<sup>118</sup> and advanced data analysis software progress has started to accelerate.<sup>119-120</sup>

LC-MS-based proteomics, mainly utilizing the bottom-up strategy, has been widely-used for functional protein studies (e.g. enzymes and drug targets), in which, function validation is always a major challenge.<sup>121,122</sup> Herein, a variety of proteomics strategies have been developed for “active” proteins characterization. Activity-based protein profiling (ABPP) represents one of the most powerful chemical proteomic strategies (Figure 1-4). In principle, the approach makes use of an activity-based probe made of by three major components: 1) an electrophilic group that covalently binds to the nucleophilic residues within the active site of an enzyme, 2) a linker that connects the reactive group with a functional tag (e.g. biotin), and 3) the tag that enables the enrichment of the probe-labeled proteins through a biochemical assay (e.g. biotin-streptavidin interaction). Dependent on the selectivity of the probe, a specific enzyme or enzyme class can be directly labeled *in situ* within a complex biological system (e.g. live cells) and enriched selectively followed by the identification and quantification of the target proteins of interest through high-resolution LC-MS/MS analysis. Since the first example of ABPP targeting esterases by Ostrowski and Barnard with radioactive probe in 1961,<sup>123</sup> many activity-based probes have been developed targeting different classes of enzymes, including serine-,<sup>124</sup> cysteine-,<sup>125</sup> aspartyl-,<sup>126</sup> and metallo-<sup>127</sup> hydrolases, kinases,<sup>128</sup> glycosidases,<sup>129</sup> histone deacetylases,<sup>130</sup> and oxidoreductases.<sup>131</sup> Despite of the robustness of ABPP, the dependence on a chemical probe somehow limits its application scope.





**Figure 1-4.** Schematic of the activity-based protein profiling (ABPP)

Several ligand label-free proteomic methods have been developed for functional protein characterization (e.g. enzyme and drug target), by monitoring protein property changes upon ligand binding. The shared principle is that small molecule-protein complexes become more resistant to various external stresses, like protease, oxidation or heat, as the ligand binding stabilizes the protein structure.<sup>132</sup> Drug affinity responsive target stability (DARTS) characterizes the protein targets by separating the protease-resistant drug-treated proteins from others in complicated cell lysate samples using the SDS-PAGE analysis, followed by in-gel proteomic identification. This approach has been applied to and identified several drug targets,<sup>133</sup> yet is limited by the sensitivity of in-gel based proteomics analysis for low abundance proteins identification. Most recently, several proteomics-based approaches, that can globally map the interaction landscape of drugs or metabolites, have been reported. Stability of proteins from rates of oxidation (SPROX) identifies the bonded proteins by reporting ligand-induced thermodynamic changes, which requires the detection and quantification of the methionine-containing peptides.<sup>134</sup> LiP-small molecule mapping (LiP-SMap) can identify protein-small molecule interactions via measuring ligand-induced protease susceptibility, through comparing the conformatypic peptide intensities among different samples that are treated by a drug with different concentrations. Similar to SPROX,

LiP-SMap characterizes the bounded proteins through measuring the intensity change of the peptide which is considered to be the binding site.<sup>135</sup> Thermal proteome profiling (TPP) is one of the representative unbiased small molecule-protein interaction mapping approaches that measures the thermal stabilization of proteins. It differentiates the binding proteins through comparing the melting curves of the drug-treated and control samples under 10 different temperature conditions.<sup>136-137</sup> The method has been widely used recently for drug target profiling.<sup>138-141</sup>

As an essential analytical tool, LC-MS-based proteomics has witnessed rapid developments and successful applications, over the past two decades, especially in profiling and quantitative analysis of proteins and their proteoforms. Nevertheless, functional protein characterization or profiling in complex samples remains a major challenge in proteomics field.

#### **1.4 Perspectives and Future Opportunities**

Identification and quantification of bioactive metabolites (e.g. natural products) and functional proteins (e.g. enzymes, drug targets) are critical in many chemobiological fields, such as drug lead discovery, drug metabolism study, chemical verification of enzymes, disease diagnosis, food quality control, and so on. The early discovery and investigation of these bioactive molecules often rely on their purification, which is time-consuming, labor-intensive, and low throughput. With the ever-emerging new advanced technologies in the past several decades, methodologies and techniques have evolved dramatically for the discovery and characterization of new bioactive molecules.

Mass spectrometry provides incomparable speed, accuracy, and sensitivity, making it the most popular tool for large-scale identification, quantification, and activity analysis of the mole-

cule of interest in complex samples, especially when coupling with modern LC separation technologies (e.g. HPLC and UHPLC). Some LC-MS-based metabolomic and proteomic platforms, such as Molecular Networking, ABPP, TPP, and many others, have been developed and extensively expanded in the past four decades. These approaches enable rapid and efficient detection, identification, and characterization of bioactive molecules from highly complex biological systems through the combination of multiple disciplines including synthetic chemistry, biochemistry, analytical chemistry, and bioinformatics. Nevertheless, these approaches often have limitations, such as the difficulty in obtaining new probes, the lack of compatibility with different biological systems, and the insufficient informatic tools for accurate structure and function characterizations. With the continuous improvement of current technologies, methodologies, informatic tools, and analytical techniques, new platforms and approaches will continue to emerge for bioactive metabolite and functional protein characterization from complex biological systems with unprecedented accuracy, sensitivity, and throughput.

### Acronym list of chapter 1

proton nuclear magnetic resonance	<sup>1</sup> H NMR
activity-based protein profiling	ABPP
anion exchange	AEX
atmospheric pressure chemical ionization	APCI
atmospheric pressure photoionization	APPI
charged aerosol detector	CAD
capillary electrophoresis	CE
collision-induced dissociation	CID
capillary isoelectric focusing	CIEF
drug affinity responsive target stability	DARTS
Dictionary of Natural Products	DNP
electron capture dissociation	ECD
evaporative light scattering detector	ELSD
electrospray	ES
electrospray ionization	ESI
fast atom bombardment	FAB

Fourier transform ion cyclotron resonance	FTICR
hepatocellular carcinoma	HCC
higher-energy collisional dissociation	HCD
hydrophobic interaction chromatography	HIC
hydrophilic interaction liquid chromatography	HILIC
Human Metabolome Database	HMDB
high-performance liquid chromatography	HPLC
ion-exchange chromatography	IC
Isotope-Coded Affinity Tag	ICAT
Isotope Coded Protein Labeling	ICPL
ion mobility spectrometry	IMS
Isobaric Tags for Relative and Absolute Quantification	iTRAQ
liquid chromatography	LC
liquid chromatography-mass spectrometry	LC-MS
LiP-small molecule mapping	LiP-SMap
liquid-liquid extraction	LLE
mass-to-charge ratio	m/z
matrix-assisted laser desorption ionization	MALDI
multiple reaction monitoring	MRM
mass spectrometry	MS
precursor ions or parent ions	MS1
product or daughter ions	MS2
natural products	NPs
principal component analysis	PCA
photodiode array	PDA
partial least squares discriminant analysis	PLS-DA
parallel reaction monitoring	PRM
post-translational modifications	PTMs
reversed-phase chromatography	RPC
strong cation exchange	SCX
size-exclusion chromatography	SEC
supercritical fluid chromatography	SFC
Stable Isotope Labeling by Amino Acids in Cell Culture	SILAC
solid-liquid extraction	SLE
solid phase extraction	SPE
stability of proteins from rates of oxidation	SPROX
selected reaction monitoring	SRM
tandem mass tag	TMT
time-of-flight	TOF
thermal proteome profiling	TPP
ultrahigh-pressure liquid chromatography	UHPLC
UV photodissociation	UVPD

## Chapter 2: Chapter Overviews

### 2.1 Overview of Graduate Research and Accomplishments

As a graduate student in the division of analytical chemistry, my research mainly focused on LC-MS/MS-based bioanalytical method development, validation, and application to detect and identify bioactive molecules (e.g. potential drug leads, functional proteins) in complex biological samples (e.g. metabolite extracts, proteomes). Through 5+ years training, I was well-trained for the whole pipelines of metabolomics (e.g. identification, quantitation), and quantitative proteomics (e.g. label-free, isotopic labeling), including: 1) sample handling/preparation, such as metabolome and proteome extraction from plasma, cancer cell lines, bacteria, fungi through homogenization, sonication or bead-beating approaches, some of which were operated in BSL-2 labs; 2) hands-on HPLC/UHPLC/LC-MS operation experience (e.g. HPLC, UHPLC with normal/micro/nano-flow rate, and MS platform of Orbitrap Fusion, Q-Executive, Quadruple, IMS-QTOF, etc.), in charge of routine calibration, maintenance, and troubleshooting as an instrument manager; and 3) data processing and interpretation using either commercial or public statistical software (e.g. MSGF+, Skyline, MZmine, molecular networking). Moreover, I am adept at new bioassay development, such as high throughput enzymatic activity measurement (e.g. biomass-degrading enzymes, kinases) and DNA/small molecules binding assays. In addition, my research work has led to multiple publications and research awards, including a Lloyd E. Swearingen Scholarship for recognition of the outstanding graduate research in the Department of Chemistry & Biochemistry and 3 travel grants to international conferences from the University of Oklahoma. I believe that my unique research experience has distinguished myself from other analytical or bioanalytical chemists.

In addition to these laboratory techniques, I have also gained good scientific reporting and communication skills. I have been actively presenting my research work in conferences and seminars. As the presenting author, I had 2 oral and 5 poster presentations at several international conferences, including ASMS, USHUPO, ISCC, and ASP. Through my experience in new method development, I have gained various abilities, including literature-reviewing, critical-thinking, independent problem-solving, multi-tasking, keen attention to detail, accuracy for data recording and organization, and so on.

## **2.2 Chapter 3. Lickety-Split Ligand-Affinity-Based Molecular Angling System (LLAMAS): A Strategy for Detecting and Dereplicating DNA-Binding Biomolecules from Complex Natural Product Mixtures**

In Chapter 3, I present the development and application of a Lickety-Split Ligand-Affinity-Based Molecular Angling System (LLAMAS): an ultrafiltration-LC-PDA-MS/MS-based DNA binding assay, coupled with modern dereplication tools (i.e., GNPS, and DNP), for detecting and dereplicating DNA-binding molecules from complex natural product extracts. The LLAMAS was developed using eight known small-molecule DNA binders and then validated by the successful identification of three DNA intercalators from *Streptomyces antibioticus*. The approach was then further optimized to 96-well plate-based approach to accommodate the required throughput in new therapeutic natural product discovery pipeline. The well-developed high throughput angling system was then successfully applied into the screening of DNA binding molecules from a mini herbal extract library from our lab and a relatively larger plant extract collection from National Cancer Institute (NCI).

### **2.3 Chapter 4. Finding Biomass-degrading Enzymes Through an Activity-Related Quantitative Protein Profiling Platform (ACPP)**

In Chapter 4, the development and application of the Activity-Related Quantitative Protein Profiling Platform (ACPP) is discussed. Specifically, the ACPP can be applied to identify functional proteins (e.g. biomass-degrading enzymes) from complex protein mixtures (e.g. fungal secretome) through cross-correlating protein-level enzymatic activity patterns with protein elution profiles obtained from the “native” LC prefractionation and quantitative proteomics analysis. The platform was developed by spiking a standard starch hydrolysis enzyme, 1,4- $\alpha$ -glucosidase, in the *Escherichia coli* lysate. It was then successfully applied to the identification of two types of active biomass-degrading enzymes (e.g. starch hydrolysis enzymes and cellulose hydrolysis enzymes) from lab-cultured *Aspergillus niger* secretome.

# **Chapter 3: Lickety-Split Ligand-Affinity-Based Molecular Angling System (LLAMAS): A Strategy for Detecting and Dereplicating DNA-Binding Biomolecules from Complex Natural Product Mixtures**

## **3.1 Introduction**

Natural products are an important source of therapeutic drug leads owing to the incredible diversity of their structures and bioactivities.<sup>60, 142-143</sup> Many of the successful stories to emerge from the field of natural-product drug-development research are legendary: the discovery of the antibiotic penicillin in 1928,<sup>144</sup> the approval of anticancer drug taxol/paclitaxel by the US FDA in 1992,<sup>145</sup> the discovery of the antimalarial drug artemisinin which won the 2015 Nobel Prize in Physiology or Medicine,<sup>146</sup> and more.<sup>147</sup> These milestones in drug discovery established natural products as an unparalleled resource for identifying therapeutically useful compounds to combat a wide range of diseases.

Strikingly, a majority of the iconic natural products that are used as medicines were discovered using the well-established, yet powerful technique known as bioassay-guided purification.<sup>61, 148</sup> This approach relies on subjecting mixtures of compounds (e.g. extracts and fractions) to iterative steps of fractionation and biological testing with the underlying strategy aimed at reducing each sample's chemical complexity until a single bioactive compound or group of bioactive substances is secured. This method is effective, intuitive, and offers tremendous rigor as researchers parse complex natural product mixtures; however, it has also been criticized for some real and perceived weaknesses: the process is somewhat slower compared to other library screening approaches, and it requires researchers to carefully track and dereplicate bioactive com-



pounds throughout the purification process.<sup>78</sup> A variety of supporting techniques have been reported to enhance bioassay-guided-purification strategies, but few fundamental changes have occurred to this central dogma of natural products drug discovery in over a century of use.<sup>149</sup>

The longevity of bioassay-guided-purification in the field of natural product drug discovery speaks volumes to its power and efficiency; undoubtedly, this technique will remain a mainstay of the field for years to come. Yet natural product drug discovery has also witnessed many extraordinary technological advances such as new tools and data resources that have the power to disrupt traditional practices like bioassay-guided-purification.<sup>150,151,152</sup> Our lab has sought to explore these emerging technologies not just as modifying agent used within existing drug discovery frameworks, but rather as potentially transformative approaches to work in parallel, or some cases even replace, existing paradigm of bioassay-guided-purification.

One area we see tremendous promise for enhancing natural-product drug discovery is based on the concept of ligand fishing. Although this approach is not new, it is certainly one that is now poised to take full advantage of the wealth of analytical tools and knowledge-based resources that have become available to natural products researchers. What makes this method so attractive as an alternative to classical bioassay-guided purification is that it allows a researcher to condense multiple rounds of bioassays and purification into a single step. This is because ligand fishing turns the biological target into both the subject of the assay, as well as a pseudo-sorbent for retaining compounds of interest.

While a detailed review of ligand fishing methods is beyond the scope of this discussion, we would like to highlight some selected cases from the natural products literature where representative technologies have been used to identify putative bioactive substances from compound

mixtures.<sup>152</sup> In principle, ligand fishing approaches enable enrichment and separation of the bioactive molecules from the inactive matrix based on their binding characteristics to the corresponding biological targets (e.g. proteins, DNAs, and RNAs, or other), through various techniques, such as immunoprecipitation, affinity chromatography, equilibrium dialysis, ultrafiltration, and so on. Fu et al. discovered several water-soluble salvianolic acid analogues as xanthine oxidase inhibitors from *Radix Salviae Miltiorrhizae* using size exclusion affinity chromatography.<sup>153</sup> The ligand fishing approach based on ultrafiltration was particularly successful for identifying enzyme inhibitors from complex medicinal plant extracts. Some representative examples include the discovery of: inhibitors of *Plasmodium falciparum* thioredoxin reductase (PfTrxR) from the MEGx® collection of AnalytiCon Discovery containing 133 structurally diverse natural compounds,<sup>154</sup> phenolic metabolites as mitochondria-targeted bioactive constituents from herbal medicines,<sup>155</sup> quercetin-3-*O*-rhamnoside and protocatechuic acid from *Kadsura longipedunculata* as  $\alpha$ -amylase inhibitors,<sup>156</sup> C-glycosylflavones as  $\alpha$ -glucosidase inhibitors from hawthorn leaf flavonoids extract,<sup>157</sup> and eupatilin from *Artemisia argyi* as a selective PPAR $\alpha$  agonist.<sup>158</sup> A relatively newer approach based on magnetic microbead affinity selection screening was invented by the van Breemen lab in 2008 to identify three estrogens from the herbal extracts of *Trifolium pratense* L. (red clover) and *Humulus lupulus* L. (hops) as inhibitors of the estrogen receptor<sup>159</sup>. In a later study, the similar approach was applied to identify quercitrin as an inhibitor of 5-lipoxygenase<sup>160</sup> from an herbal extract of *Proserpinaca palustris*.

For simplicity, these approaches can be divided into two major categories: those that involve binding or immobilization of the biological target to a solid substrate and those that forego the immobilization requirement. Among the techniques dependent on target immobilization, magnetic beads have been used to identify molecules that bind to proteins/enzymes from natural

product extracts.<sup>161-162</sup> A functionally similar approach has been achieved wherein the protein/enzyme of interest is bound to the surface of a capillary or pre-capillary chamber. After incubating the bound target with a compound mixture, electrophoresis was used to detect protein-binding substances based on their increased retention. This method was used to identify putative inhibitors of epidermal growth factor receptor via capillary electrophoresis coupled to LC-MS.<sup>163</sup> Unfortunately, both of these methods suffer from the challenges surrounding the need to effectively perform the immobilization of a target substrate to a bead, capillary, or other inert surface.

Ligand fishing techniques that do not require target immobilization offer certain advantages over target-binding-based approaches due to a combination of enhanced assay simplicity, the capacity to retain a biological target's native form, and the ability to readily use a wider range of biological targets. Dialysis-based methods represents one of these types of strategies in which a semi-permeable membrane is employed to afford the enrichment of putative ligands from compound mixtures via molecular diffusion. Over time, a dynamic equilibrium is established, wherein a molecule with an affinity for a particular biological target, e.g., will become concentrated with the target of interest. In theory, this approach should provide a mechanism to select for compounds from complex mixtures; however, in practice, it faces some problematic challenges. Namely, this technique is hampered by its relatively low-throughput due to the lengthy periods required to establish dynamic equilibrium. Nevertheless, the Liu group reported the use of an equilibrium dialysis system that employed LC-MS detection to successfully characterize seven known liposome binders from *Panax ginseng* extract.<sup>164</sup> Similar to this methods, ultrafiltration coupled with LC-MS analysis is a powerful approach to detect potentially bioactive substances from complex samples. This technique requires little or no sample pretreatment and can be coupled to a variety of analytical readout devices. Recently, Li and Yang reported an

strategy using ultrafiltration with LC-MS in conjunction with in silico molecular docking to discover enzyme inhibitors from traditional plant-based medicines.<sup>165</sup>

One of the cellular components that continues to be an attractive, druggable target is DNA. For decades, molecules that bind to DNA have served as powerful and clinically effective weapons for combating several types of cancer. Examples of DNA-binding agents used in cancer chemotherapy include mitomycin C,<sup>166</sup> cisplatin,<sup>167</sup> and actinomycin D (**9**);<sup>168</sup> however, their success is somewhat limited due to dose-limiting side effects.<sup>169-170</sup> Despite these success, the discovery of new therapeutically-useful DNA binding molecules remains a challenging proposition in part because of DNA's limited range of recurring binding motifs.<sup>171</sup> Nevertheless, a promising report by Liu and Wan revealed that a triplex-DNA-binding assay was capable of detecting compounds from natural plant extracts using a DNA-conjugated agarose-bead-based baiting technique that was used in combination with LC-MS detection.<sup>172</sup> Unfortunately, this method still required the DNA to be immobilized, which presents certain challenges in high-throughput assay situations.

Our goal was to develop an efficient ligand fishing assay for the detection and identification of DNA-binding agents from natural product mixtures, which can be easily adapted for the high throughput screening required in common natural product discovery pipeline. As part of that goal, we sought to use non-immobilized DNA that would allow the assay to be readily converted for use with different DNA sources, sequences, and targets. In this report, we describe our efforts to implement a natural product drug discovery pipeline using an ultrafiltration-based assay system linked with LC-PDA-MS/MS to seamlessly detect, dereplicate, and in some cases identify compounds from complex mixtures of natural products (e.g. total organic extracts). For higher throughput, the developed approach was then optimized to 96-well plate-based screening.

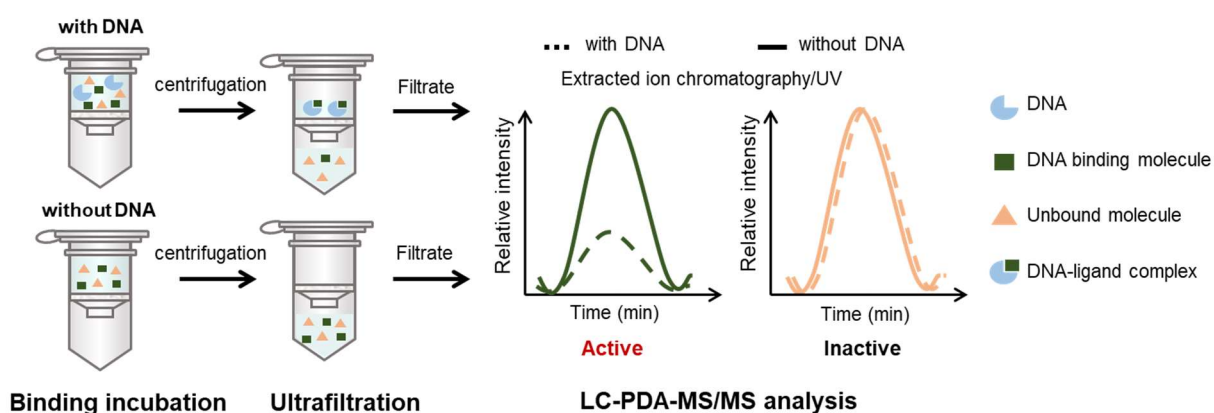
With the optimized assay, 62 herbal extracts from our lab and 332 plant samples obtained from NCI were screened, with 3 unique hits identified to contain DNA binding molecules.

## **3.2 Results and discussion**

### **3.2.1 Design and Implementation of an Ultrafiltration-LC-PDA-MS/MS-Based DNA-Binding Assay**

A ligand fishing strategy was devised to identify DNA-binding agents from complex mixtures of natural products (Figure 3-1). While ligand fishing techniques are regularly exploited in the field of chemical biology, we found comparatively fewer cases of this methodological approach being used for the purpose of chemical detection in natural product drug discovery.<sup>152,173</sup> This was surprising to us, since in theory, ligand fishing approaches have the power to combine the individual steps of a biological assay with analytical chemical detection to create a seamless process that in theory, is capable of shortening the iterative cycles of traditional bioassay-guided compound identification. In our adaptation of ligand fishing, the plan consisted of four successive stages: (1) an incubation phase to afford binding of compounds with their DNA targets, (2) ultrafiltration to separate the ligand-bound-DNA complex from unbound small molecules, (3) untargeted LC-MS analysis of the filtrates to detect candidate DNA-binding molecules, and (4) employment of natural-product data resources (e.g. GNPS, SciFinder, DNP, and others) to both dereplicate and guide efforts toward the identification of putative DNA-binding molecules. By comparing the filtrates of extracts incubated with DNA versus control samples that were processed without DNA, it was reasoned that compounds bound or otherwise associated with DNA would be revealed based on their differential abundance in experimental versus control filtrates. The filtrates from the experimental (with DNA) and control (without DNA) conditions could

then be analyzed directly using an ultra-high-performance-liquid-chromatography (UHPLC or simply LC) system equipped with a photo-diode-array (PDA) detector that was coupled to an ion trap mass spectrometer with ion fragmentation capabilities (MS/MS). We rationalized that by employing the orthogonal detection capabilities of LC-PDA-MS/MS (i.e., UV and EIC traces), it would provide a sensitive analytical platform that could handle a broad range of chemical scaffolds such as those found among natural products. Furthermore, if the system were operated within the linear dynamic range of putative binding agents (Figure S2, Appendix), the peak areas for DNA-binding molecules would measurably decrease within the experimental group, whereas the peak areas for non-DNA binding molecules would remain unchanged.



**Figure 3-1.** Overview of a strategy designed to detect DNA-binding molecules from chemical mixtures using an ultrafiltration assay coupled to LC-PDA-MS/MS. Using this approach, putative DNA binding molecules would be detected by comparing the filtrates from samples that were incubated with (experimental group) or without (control group) DNA.

Many experimental parameters were considered during the development phase of the project and optimized for our experimental; the reasoning behind some of the experimental design decisions merit further discussion. Bulk salmon-sperm DNA was used for our studies since it





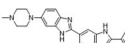
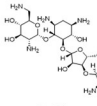


represented an affordable source of reasonably-sized, intact DNA [double helix fragments consisting on average of  $\approx 2,000$  bp (1,300 kDa)] that would be widely available to other labs wanting to adopt our protocols. After considering several types of ultrafiltration membranes, a modified polyethersulfone membrane that offered a 100 kDa cutoff was selected since it presented an advantageous suite of properties (compatible with a range of elution solvents and reasonable inertness). Moreover, elution could be conducted using unextraordinary centrifugation conditions (i.e.,  $5,000 \times g$ ) making this technique adaptable to many lab settings. A variety of assay incubation conditions were also evaluated as we searched for a solution that would allow for the DNA to retain a natural double helix structure, enable the solubilization of wide range of natural products, and minimize the number of weakly bound compounds in favor of molecules that exhibited stronger DNA binding interactions. This led to the identification of a modified glycerol-containing Tris-EDTA buffer combined with 33% by volume MeOH. The buffer served as the incubation solution and served in the wash step to remove unbound molecules. In our hands, this solution enabled a wide range of compounds (including many hydrophobic substances) to remain in solution during the incubation period.

### **3.2.1.1 Testing Assorted DNA-Binding Agents**

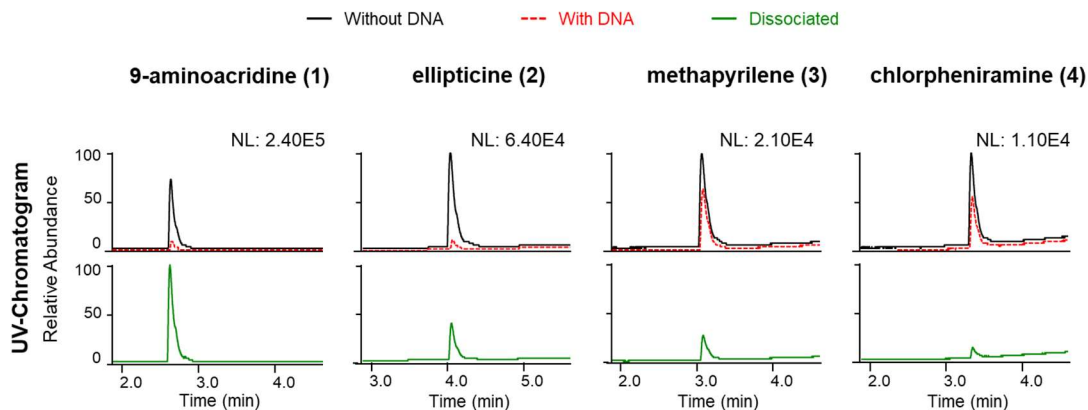
Eight compounds known to interact with DNA (Table 3-1) were selected to test the ligand fishing system. Those compounds were chosen based on several criteria including the inclusion of diverse structural features and chemical properties, coverage of different DNA binding mechanisms, and non-equivalent sensitivities to detection by UV and MS instrumentation. Two assay endpoints were examined (Figures 3-2 and S3, Appendix). The first endpoint involved determining whether a putative DNA-binding molecule was present in the eluent obtained from the

ultrafiltration wash step based on semi-quantitative comparisons of its relative concentrations in samples incubated with and without DNA (Figure 3-2, upper panels). The second endpoint relied on analyzing the profile of substances obtained from an organic solvent rinse step (MeOH spiked with 0.1% formic acid) used to disrupt small-molecule binding interactions with DNA after the buffer wash (Figure 3-2, lower panels). While both methods were deemed to be informative and provide complementary information [e.g. distinguishing between covalent versus non-covalent DNA binders (Figure 3-3)], we conclude the first method alone was sufficient to perform routine sample screening. Thus, we determined that our ligand fishing system could detect molecules that exhibited different DNA binding mechanisms [i.e., intercalators (**1-4**), groove binders (**5-6**), and covalent binders (**7-8**)], and covered a wide range of structural motifs including molecules that were difficult to ionize under standard ESIMS conditions [i.e., compound **8**].

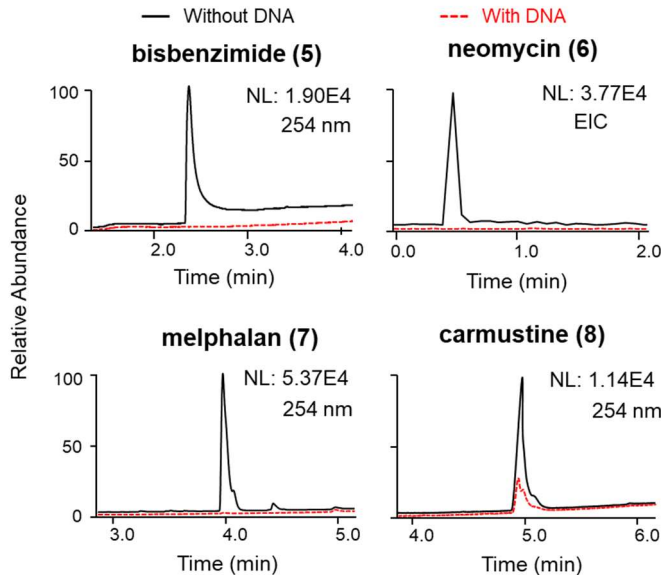
**Table 3-1.** Information of 8 known DNA binders

Name	Structure	Binding mechanism	Solubility in aqueous solution	Distinct UV absorbance	[M+H] <sup>+</sup>
9-aminoacridine ( <b>1</b> )		intercalator	poor	yes	195.09
ellipticine ( <b>2</b> )		intercalator	poor	yes	247.12
methapyrilene ( <b>3</b> )		intercalator	poor	yes	262.14
chlorpheniramine ( <b>4</b> )		intercalator	good	yes	275.13
bisbenzimidazole ( <b>5</b> )		groove binder	poor	yes	425.21
neomycin ( <b>6</b> )		groove binder	good	N/A	615.32
melphalan ( <b>7</b> )		covalent binder	poor	yes	305.08
carmustine ( <b>8</b> )		covalent binder	poor	yes	N/A





**Figure 3-2.** Methods for the detection of intercalating compounds **1-4** in the DNA-binding assay. The UV chromatograms ( $\lambda$  254 nm) show the peak areas of the filtrate for the control group samples incubated without DNA (upper panels, black solid line traces) compared to experimental group samples that were incubated with DNA (upper panels, red hashed line traces). The compounds bound with the DNA were subsequently dissociated (lower panels, green solid line traces) with MeOH containing 0.1% formic acid. (NL: normalized intensity)

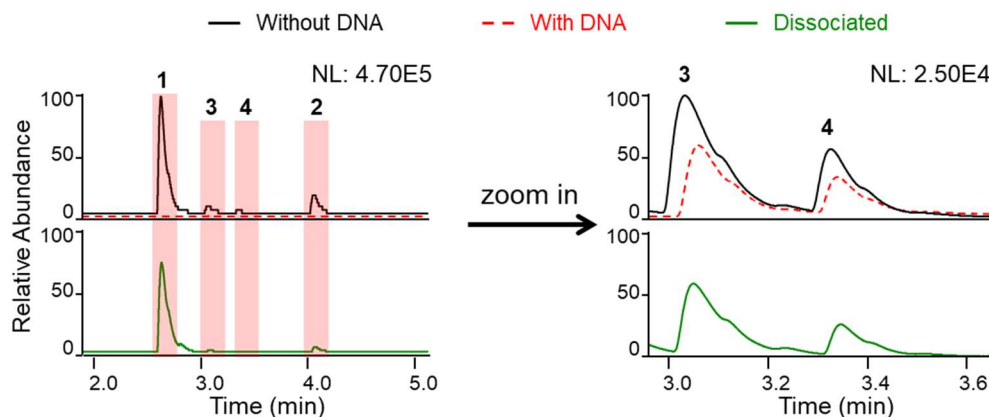


**Figure 3-3.** Detection of groove-binding agents **5-6** and covalent-binding compounds **7-8** in the DNA-binding assay. Compounds **5**, **7**, and **8** were observed by UV ( $\lambda$  254 nm) detection, whereas **6**, which lacks a clearly detectable UV chromophore was monitored using the EIC trace from the mass spectrometry.

ter. Individual plots show the peak areas for the compounds from the filtrate for the control group (incubated without DNA, black solid line traces) superimposed on the traces recorded for the experimental samples (incubated with DNA, red hashed line traces). While covalent-binding compounds **7** and **8** unsurprisingly were not dissociated from the DNA, the groove-binding compounds **5** and **6** were also proved difficult to recover using the dissociation solvent (data not shown). (NL: normalized intensity; EIC: extracted-ion chromatogram)

### 3.2.1.2 Detection of DNA-Binding Molecules in Mixtures and Complex Matrices

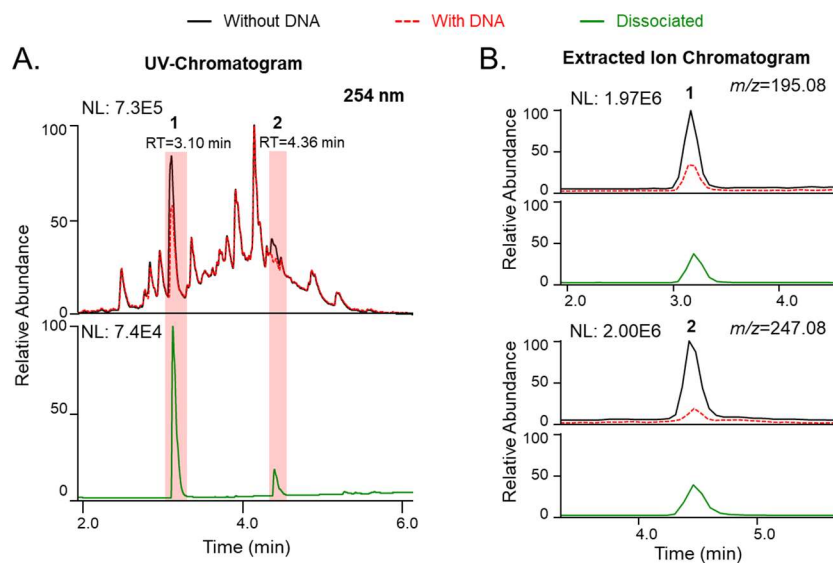
To establish the outcome when multiple types of compounds competed for the same or similar DNA-binding sites, we prepare a mixture of four DNA-intercalating agents (final assay concentration 51  $\mu\text{M}$  of **1**, 81  $\mu\text{M}$  of **2**, 67  $\mu\text{M}$  of **3**, and 51  $\mu\text{M}$  of **4**) for testing. We rationalized that many of the metabolites encountered in natural product extracts occur as mixtures; therefore, competition for binding sites was likely to occur during sample screening. The results from our tests indicated that such mixtures would likely not be problematic since all compounds were readily detected by LC-PDA-MS/MS (Figure 3-4).



**Figure 3-4.** Monitoring the outcome for four DNA intercalators (**1-4**) tested as a mixture in the DNA-binding assay. The upper UV traces ( $\lambda$  254 nm) show the results for compounds incubated with (red hashed lines) or without (black solid lines) DNA whereas the lower panel (green solid lines) show the UV

traces following compound dissociation using MeOH containing 0.1% formic acid. (NL: normalized intensity)

Next, we sought to test how the assay system would perform when a DNA binding compound incorporated into a complex matrix. We employed what was perceived to be a potential “worst-case” scenario that consisted of an organic extract prepared from soil (the top  $\approx 7$  cm of material collected from a low-lying hardwood forest plot that supported lush herbaceous plant growth). The soil extract contained  $>1,000$  potential chemical features found in a wide range of concentrations.<sup>174</sup> Two DNA intercalators, 9-aminoacridine (**1**) and ellipticine (**2**), were spiked into the soil extract in low 1:600 (w:w) amounts to test the assay system’s robustness. Analysis revealed that even under those challenging conditions, both compounds were readily detected (Figure 3-5) affording confidence that the assay was capable of handling the level of chemical complexity present in most natural product crude extracts.



**Figure 3-5.** Detection of DNA binding agents incorporated into a complex soil extract. DNA intercalators **1** and **2** were added (1:600, w:w) into an organic extract prepared from soil. (A) The UV chromatograms ( $\lambda$  254 nm) reveal that retention of the test compounds was observable upon comparison of the samples incubated with (red hashed lines) versus without (black solid lines) DNA. Additionally, compounds **1** and

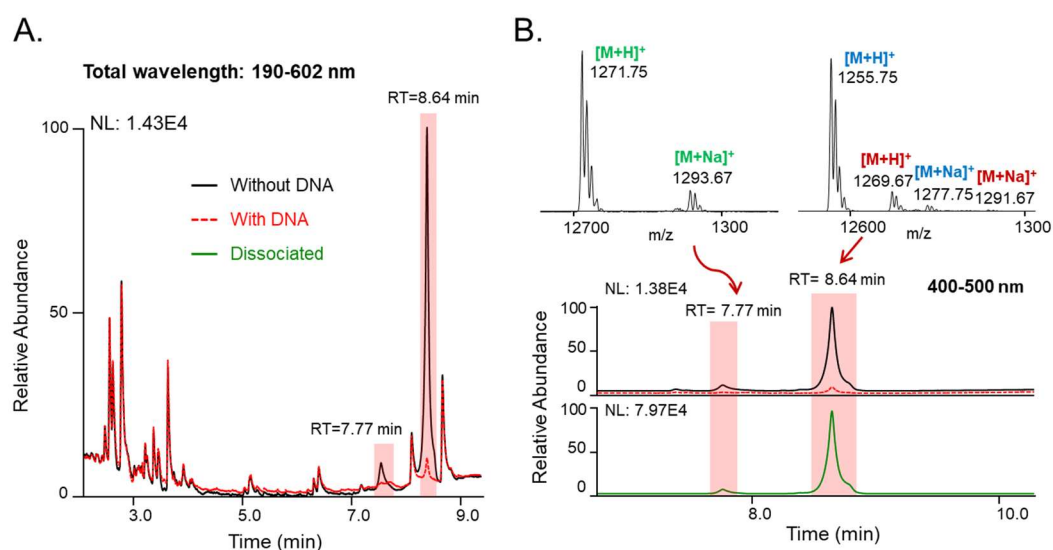
**2** were readily found following their dissociation from the DNA using MeOH containing 0.1% formic acid (green solid line shown in the lower panel). (B) Alternatively, the extracted ion chromatograms demonstrated mass spectrometry could also serve as was a suitable tool for detection in a complex substrate. (NL: normalized intensity; RT: retention time)

At this juncture, we concluded that our system of detecting and analyzing putative DNA-binding compounds was sufficiently refined and repeatable that it was ready for testing natural product extracts. We were encouraged by how quickly single samples could be processed giving us reason to believe that this approach could be further engineered to offer a reasonable level of throughput for screening purposes. This led to some debate within our lab what to call our “ultra-filtration-LC-PDA-MS/MS-based DNA-binding assay and compound identification process,” since its analytical capabilities and potential screening applications went beyond traditional ligand fishing methods. Internally, we moved to adopt the amusing acronym “LLAMAS 1.0,” which stood for Lickety-split Ligand-Affinity-based Molecular Angling System version 1.0 as a amusing term intended to represent the unique processes built into our all-inclusive platform.

### **3.2.1.3 Testing LLAMAS 1.0 for the Identification of DNA-Binding Natural Products in a Microbial Extract**

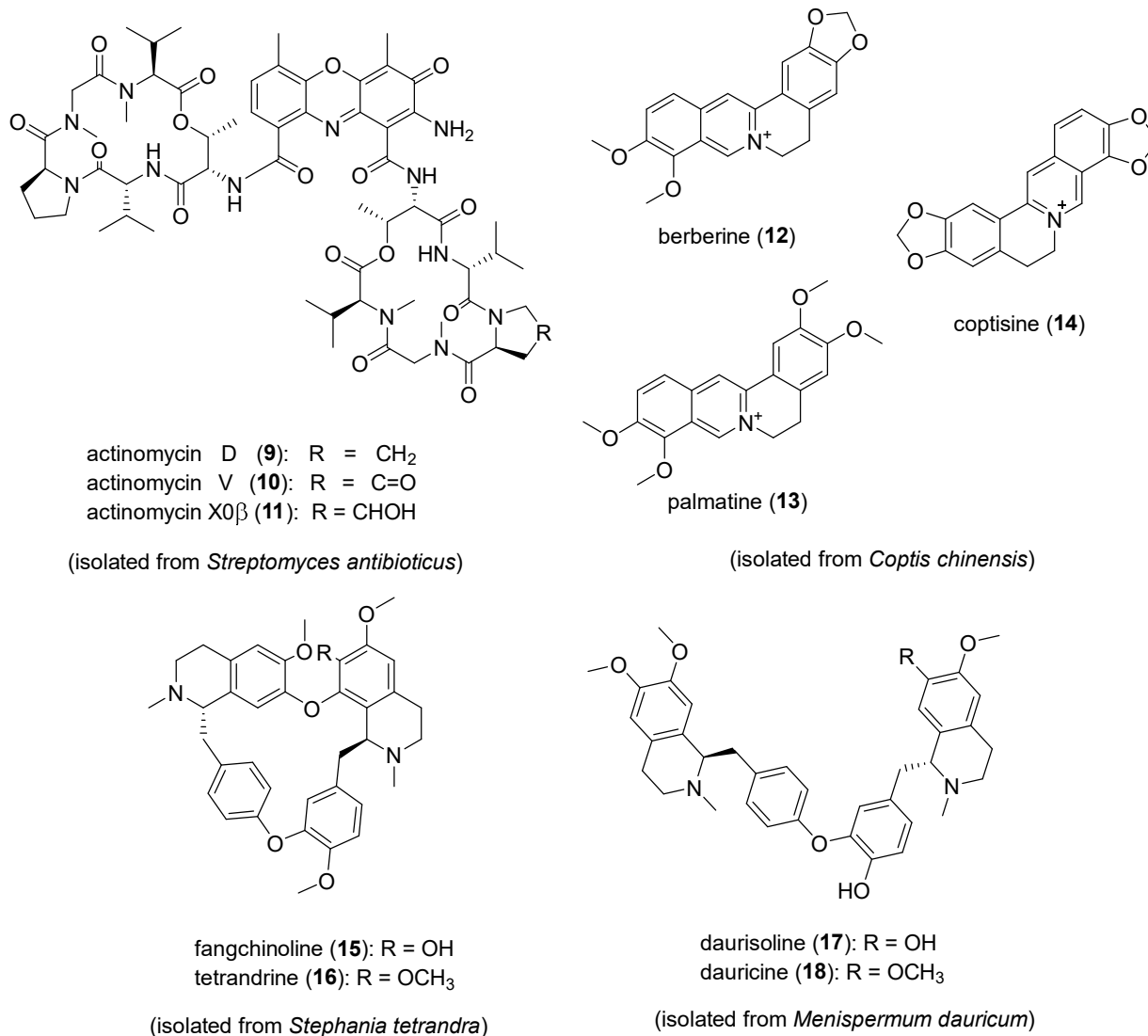
Actinomycin D (**9**) and its analogues are DNA-intercalating agents found in several *Streptomyces* spp.<sup>175-177</sup> We prepared a sample containing the ethyl-acetate-soluble components from a *Streptomyces antibioticus* (ATCC14888) culture and analyzed the resulting extract using LLAMAS 1.0. The analysis resulted in the detection of multiple putative DNA binding candidates including a major UV-active peak (RT 8.64 min) that afforded two mass features ( $m/z$  1255.75 and 1269.67), as well as a minor UV-active peak (RT 7.77 min) offering a single mass

feature ( $m/z$  of 1271.75) (Figure 3-6). The resulting MS data were analyzed using the GNPS open-source cheminformatic platform leading to the provisional identification of actinomycin D (**9**) (Figure S5, Appendix) along with two metabolites that were likely to be structural analogues of **9** based on their mass fragmentation data. To confirm the identities of the compounds, MS-guided semi-preparative  $C_{18}$  HPLC was used to purify the three metabolites, which were subsequently subjected to NMR and other spectroscopic tests resulting in their authentication as actinomycin D (**9**), V (**10**) and  $X_{0\beta}$  (**11**) (Figure 3-7).<sup>178</sup> Thus, the incorporation of molecular networking into the LLAMAS platform demonstrated the potential to accelerate the identification/dereplication of DNA-binding compounds from multicomponent natural product samples.



**Figure 3-6.** Identification of the DNA binding natural products actinomycins D (**9**), V (**10**), and  $X_{0\beta}$  (**11**) from *S. antibioticus*. (A) Analysis of the full PDA chromatogram ( $\lambda$  190-602 nm) revealed putative DNA-binding substances in the bacterial extract (candidate peaks are highlighted in red boxes). (B) Upon adjusting the PDA chromatogram to display a narrower range of wavelength ( $\lambda$  400-500 nm) (solid black and dashed red lines), as well as carrying out dissociation of the substances bound to the DNA (solid green line, lower panel), the MS data revealed the presence of three substances in the extract: **9** (RT =

8.64,  $[M+H]^+$  ion at  $m/z$  1255.75), **10** (RT = 8.64,  $[M+H]^+$  ion at  $m/z$  1269.67), and **11** (RT = 7.77,  $[M+H]^+$  ion at  $m/z$  1271.75).



**Figure 3-7.** Structures of identified DNA binding natural products (**9-18**) from bacterial and plant extracts.

## 3.2.2 LLAMAS 2.0 for 96-Well Plate-Based Assays

### 3.2.2.1 Developing LLAMAS 2.0

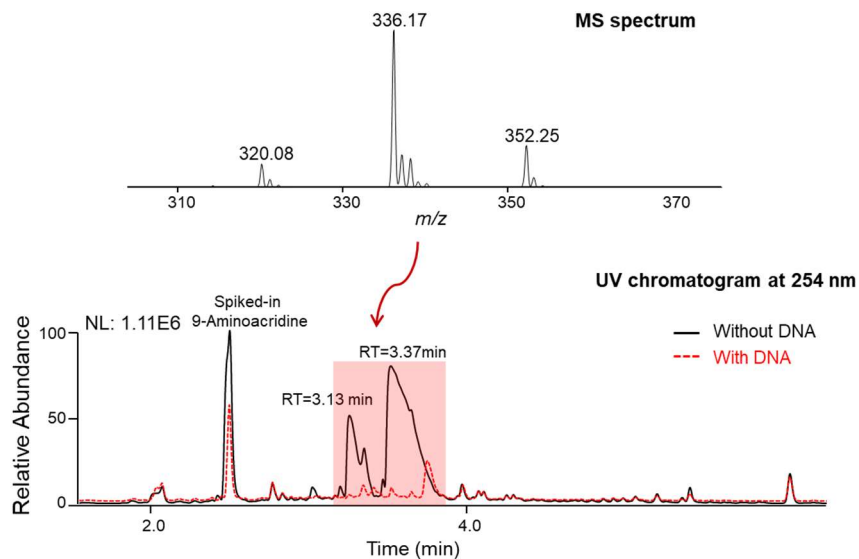
Considering the speed with which a single sample could be tested, and its DNA-binding compounds identified using LLAMAS 1.0, we speculated that substituting the filtration microtubes for a device with higher throughput potential would enhance our ability to more rapidly screen larger numbers of test substances. Therefore, we turned to evaluating a 96-well microtiter plate system that contained a 100 kDa cutoff ultrafiltration membrane, which could be used in conjunction with a vacuum manifold system to afford a significantly increased level of throughput. This approach, which was dubbed LLAMAS 2.0, was assessed using DNA binders **1-8** leading to the conclusion that all the test compounds could be detected alone (Figure S6A, Supporting Information) and in mixtures (Figure S6B, Supporting Information) using the 96-well ultrafiltration plate format. As an additional test, a mixture of wild herbaceous annual and perennial plants (unidentified assemblage of plants containing  $\approx$  15-20 dicotyledons and monocotyledons) found growing in ungrouted plots on the University of Oklahoma campus were collected and extracted creating a complex natural-product-containing extract. Compounds **1** and **4** were added to the mixed plant extract in a ratio of 1:5:250 (w:w:w of **1**:**4**:extract, respectively) and the sample tested using LLAMAS 2.0. The spiked-in DNA intercalators were readily detected employing sample sizes as small as 250  $\mu$ L in a working volume of 100  $\mu$ L (Figure S6C, Supporting Information), which was far better than the 400  $\mu$ L working volume 2-3 mg of extract needed for the microtube-based LLAMAS 1.0.

### 3.2.2.2 Testing LLAMAS 2.0 Using a Collection of Herbal-Supplement Extracts

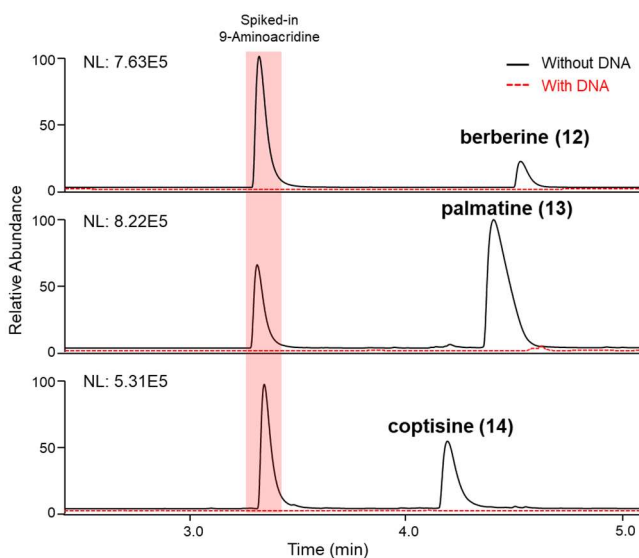
Over the course of many years, our lab has acquired a modest set of 62 plant specimens that are used in the United States in the herbal supplements industry (Table S1, Appendix). Organic extracts were prepared from the samples and portions of each were formatted in a 96-well microtiter plate for testing using LLAMAS 2.0. The experimentalist performing the test was blinded to the identities of the extracts to provide greater rigor for the experiment. The tester classified one sample as a “hit” after it was found to contain putative DNA-binding molecules (Figure 3-8) based on comparisons of its UV chromatograms and total ion traces of sample filtrates acquired following incubation of the extract with and without DNA present. The hit sample was determined to contain three UV-absorbing peaks that exhibited  $m/z$  values of 320.08, 336.17, and 352.25. The MS<sup>2</sup> fragmentation data for the three analytes were submitted to the GNPS platform, which generated a strong match to the DNA-intercalating compound berberine (**12**), as well as two berberine analogs.<sup>179-180</sup> To confirm the identities of the natural products, semipreparative C<sub>18</sub> HPLC was used to purify the compounds, which were dereplicated by NMR analyses as **12**, palmatine (**13**), and coptisine (**14**) (Figure 3-7).<sup>181</sup> Unmasking of the plant’s identity revealed the source of **12-14** to be the roots of *Coptis chinensis* Franch., which is a well-established natural source of these metabolites. The DNA-binding activities of compounds **12-14** were confirmed (Figure 3-9) and found to be consistent with published data.<sup>179, 182-183</sup> In further support of the rigorousness and correctness of the results afforded by LLAMAS 2.0, a *post hoc* literature search was performed on the other 61 plant specimens and none were reported to contain known DNA-binding compounds. The results of this test, although limited in scope, suggested that LLAMAS 2.0 was not overly vulnerable to false-negative or false-positive results under real-world natural-product screening conditions. For transparency purposes, it must be noted



that the plant materials used in this study were received as small pieces or in powdered form negating the opportunity to conduct macro-scale visual authentication.



**Figure 3-8.** Three molecules in the crude extract of root slices of Gold Thread (*Coptis chinensis*) were found to bind with DNA. UV chromatogram at 254 nm showed two major DNA binder peaks that were eluted at 3.13 min and 3.37 min. MS spectrum revealed three  $m/z$  values at 320.08, 336.17, and 352.25 for the combined region of the three UV peaks.



**Figure 3-9.** Confirmation of the DNA binding activities of berberine (**12**), palmatine (**13**), and coptisine (**14**). UV chromatogram at 254 nm showed their peaks in experimental group (red line) disappeared, while detected in control group, indicating their DNA binding activity.

### 3.2.2.3 Applying LLAMAS 2.0 to Test a Library of Traditional Chinese Medicinal Plant

#### Extracts

The US National Cancer Institute (NCI) has assembled a library composed of 332 organic extracts prepared from plants used in Traditional Chinese Medicine (TCM), and samples are available to researchers upon written request.<sup>184</sup> This library, which offers good coverage of diverse natural product scaffolds originating from several plant families, offered an excellent opportunity for testing LLAMAS 2.0. Testing of the 332 organic extracts in duplicate yielded three samples that were identified as containing DNA-binding compounds. Upon examination of the sources of the three active extracts, we noted that one of the samples had been prepared from *C. chinensis*; the data derived from LC-MS/MS revealed that the same three DNA-binding metabolites we previously dereplicated, **12-14** (*vide supra*), were present in the new sample from the NCI TCM collection.

The two remaining active samples were detected using LLAMAS 2.0 had been prepared from *Stephania tetrandra* S. Moore and *Menispermum dauricum* DC. Focusing first on the results obtained for *S. tetrandra*, two molecular features were detected that yielded  $m/z$  values of 609.33 and 623.33 ( $[M+H]^+$  ions), and were tentatively attributed to the bisbenzylisoquinoline metabolites fangchinoline (**15**) and tetrandrine (**16**), respectively (Figure 3-7), based on their biogenic source, UV-Vis spectra, reported DNA binding activities, and MS data.<sup>185,186</sup> LC-MS data obtained from the samples using the *M. dauricum* extract supported the presence of two active compounds that were supported by the presence of  $[M+H]^+$  ions exhibiting  $m/z$  values of 611.50

and 625.42, which were preliminarily identified as daurisolone (**17**) and dauricine (**18**), respectively (Figure 3-7), based on their LC-MS data, biogenic source, and reported DNA binding activities.<sup>187,188</sup> The initial structure assignments of **15-18** were subsequently confirmed based on comparisons of their <sup>1</sup>H NMR spectra, specific rotation values, and MS features with reported data.<sup>186, 189, 190-191</sup> The DNA-binding activities of purified **15-18** were then verified using the LLAMAS 2.0 method (Figure S7, Appendix).

### 3.2.3 Conclusions

The results of these studies demonstrate that LLAMAS 2.0 is an effective platform for the detection and dereplication of DNA-binding natural products from complex chemical mixtures. While this initial application LLAMAS focused on compounds that interact with DNA, we see tremendous potential for expanding this methodology to include alternative biological targets (e.g., proteins, RNA, cellular organelles, and more). The enduring dominance of bioassay-guided fractionation in natural products discovery is a testament to its power and practicality. However, alternative approaches that take advantage of the confluence of an expanding range of readily accessible commercial materials and the ever-improving capabilities of analytical tools need to be considered as enabling components for creating alternative bioactive natural product detection measures. Whereas various forms of ligand fishing techniques have been in use for a long time, it is our opinion that they have not been appropriately exploited in the field of natural products discovery. We anticipate that LLAMAS and related approaches, which consolidate the detection and identification of biologically intriguing compounds into a single process have the potential to further enhance the natural-products-driven drug discovery efforts.

### 3.3 Materials and Methods

#### 3.3.1 General Experimental Procedures

Column chromatography was performed using silica gel and HP20SS. Preparative HPLC was processed on a SHIMADZU System equipped with LC-6AD HPLC pump, coupled to SPD-M20A PDA detector and Phenomenex Luna C18 column (21.2 × 250 mm, and 10 × 250 mm, 5 μm). Analytical and semipreparative HPLC were conducted using a Waters HPLC system with 1525 binary pump and 2998 PDA detector, Phenomenex Gemini 5 μm C<sub>18</sub> and Phenomenex Kinetex pentafluorophenyl (250 × 4.6 mm, 1 mL/min, and 250 × 10 mm, 4 mL/min, 5 μm) columns. NMR data were collected on Varian 600 MHz NMR spectrometers. Microcentrifuge tube-based ultrafiltration filters (100 kDa) were obtained from Pall Corporation (Houston, Texas, USA). Salmon sperm DNA and all other chemicals were purchased from Sigma-Aldrich (St. Louis, Missouri, USA). Distilled water was purified by a Milli-Q water purification apparatus (Millipore, Bedford, MA). All solvents were of ACS grade or better.

#### 3.3.2 Culture and extraction of *Streptomyces antibioticus*

The bacterial strain (14888) was purchased from the American Type Culture Collection (ATCC, Manassas, Virginia, USA). *Streptomyces antibioticus* was first retrieved on Yeast Malt Agar plate, containing 0.5% peptone, 0.3% yeast extract, 0.3% malt extract, 1% dextrose, and 0.15~0.2% agar (final pH 6.2 ± 0.2), then cultivated in liquid media, containing 0.5% tryptone and 0.3% yeast extract. The single colony was inoculated to 20 mL liquid medium in falcon tubes, shaking at 200 rpm, 30 °C for 24 hours for seed broth preparation, which was then added to 1.25 L each liquid media for incubation under shaking at 200 rpm, 30 °C for 6 days. Around 5

L fermentation broth was obtained and then extracted with ethyl acetate through homogenization and overnight soak for three times.

### **3.3.3 Herb extraction**

Herbal materials were purchased from Dandelion Botanical Company (Seattle, WA). 2 g dry herb sample was soaked in 200 mL MeOH for 24 hours. Supernatant was dried using rot vap and then resuspend in 400 mL ethyl acetate : H<sub>2</sub>O (v : v = 1 : 1) for partitioning. The ethyl acetate phase was collected and dried for further analysis. For Gold Thread (*Coptis chinensis*), 550 mg extract was obtained from 20 g root slices using the same protocol.

### **3.3.4 Lickety-split Ligand Affinity based Molecular Angling System (LLAMAS)**

Pure compounds were dissolved in water or cosolvent of DMSO/methanol based on their solubilities. 1 mg/mL DNA solution in 1x TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) with 15% glycerol was stocked at 4 °C before usage. 1x TE buffer with 15% glycerol was prepared and stocked at 4 °C as control buffer.

In binding incubation step of the microtube-based approach, 5~50 µg pure compounds or 1~3 mg of natural extract mixtures (water-soluble pure compounds were dissolved in 200 µL water, other compounds and complex extracts were dissolved in 200 µL MeOH) was incubated with 400 µL DNA solution in experimental group, while with 400 µL control buffer in control group, at room temperature for 30 mins with gentle shaking. The incubated sample was filtered through the ultrafiltration membrane (100 kDa) at 5,000 g, 10 °C. The filtrates were collected for LC-PDA-MS/MS analysis. In dissociation workflow shown in Figure S1 (Supporting information, Appendix), the DNA-ligand complex (leftover solution above the ultrafiltration membrane) in

experimental group was washed 3 times using 30% methanol through ultrafiltration to remove the free molecules. Next, the complex was transferred to a microcentrifuge tube for dissociation in 600  $\mu$ L 95% HPLC methanol with 1% formic acid, by rigid vortex for 20 mins at room temperature. The released ligand was then separated from denatured DNA through centrifugation using a new ultrafiltration filter at 5,000 g for 10 mins. The collected filtrate was dried through evaporation in vacuum and then resuspended in 50  $\mu$ L HPLC methanol, prior to LC-MS/MS analysis.

In high throughput LLAMAS, 1~25  $\mu$ g pure compounds or 125  $\mu$ g of complex samples in 50  $\mu$ L MeOH, 100  $\mu$ L DNA solution (experimental group) or control buffer (control group), and 0.5  $\mu$ g known DNA intercalator **1** were added into each well orderly, prior to incubation at room temperature for 30 mins with occasional shaking. During incubation, the plate was sealed with sealing film to avoid evaporation. Ultrafiltration was conducted in a MultiScreen<sup>®</sup><sub>HTS</sub> Vacuum Manifold (EMD Millipore, Billerica, MA) with the vacuum level of less than 20 in. Hg (i.e. 15-20 in. Hg). Collected filtrate was directly introduced into LC-MS/MS analysis with 10  $\mu$ L sample injection.

### **3.3.5 Untargeted LC-PDA-MS/MS analysis**

LC was performed on an Accucore Vanquish UHPLC, equipped with a (photodiode array) PDA detector, and an Accucore C18 column (1.5  $\mu$ m, 100  $\times$  2.1 mm, 0.4 mL/min.). Mobile phases applied in LC separation are H<sub>2</sub>O (A) and acetonitrile (B), with 0.1% formic acid, respectively. For pure compounds and their assay output, gradient started at 7% B for 0.5 min, linearly increased from 7% to 50% B over next 5 min, then jump to 95% B in 0.5 min, hold at 95% B for

another 0.5 min and return to 7% B in 0.5 min, lastly hold at 7% B for 2 min for column re-equilibration. For soil and microbial extract and assay output, the gradient starts at 3% B, with an extended linear gradient from 3% to 95% B in 10 min; while the separation gradient is set as 3% to 80% B in 7 min, for herbal extracts. Column temperature was maintained at 40 °C and sample compartment at 10 °C during the analysis. The flow rate was set at 0.4 mL/min. MS data was acquired on Thermo LTQ XL mass spectrometer through electro-spray ionization (ESI). Capillary settings were 270 °C, 18 V with a spray voltage of 4.5 kV. Sheath gas (N<sub>2</sub>) and auxiliary gas (N<sub>2</sub>) were set at 40 arb and 5 arb, respectively. Tube lens voltage was set at 95 V for positive-ion mode. Collisional induced dissociation (CID) was applied for MS/MS fragmentation with a normalized collision energy setting of 35%, in the data dependent acquisition mode using the five most abundant parent ions.

### **3.3.6 Dereplication and characterization of DNA binding candidates**

Assay output of known DNA binders were manually characterized using their retention time, MS, and MS/MS features. In bacterial and herbal extract analysis, DNA binding candidates were dereplicated through the Global Natural Products Social Molecular networking (GNPS) online platform, using the mgf file exported from MZmine.

### **3.3.7 Fractionation and isolation of DNA binding candidates**

All crude extracts were pre-fractionated by vacuum-liquid chromatography (HP20ss), respectively, using step gradient of MeOH–H<sub>2</sub>O (30 : 70, 50 : 50, 70 : 30, 90 : 10, and 100 : 0) as eluents.

The fourth HP20ss fraction (MeOH–H<sub>2</sub>O, 90:10) from 1.3 g extract of *Streptomyces antibioticus* contains the 3 candidates with m/z of 1255.75, 1269.67, and 1271.75. It was further fractionated by C18 preparative HPLC (gradient elution using acetonitrile–H<sub>2</sub>O with 0.1% formic acid, 58% to 80% acetonitrile in 30 min) followed by semipreparative HPLC using a pentafluorophenyl column (isocratic Acetonitrile–H<sub>2</sub>O with 0.1% trifluoroacetic acid, 52 : 48), obtaining 3 pure compounds: actinomycin D (**9**, 14.9 mg), V (**10**, 10.0 mg), and X<sub>0β</sub> (**11**, 9.8 mg). The <sup>1</sup>H and <sup>13</sup>C NMR data and spectra of the isolated compounds were included in supporting information.

The second and fourth HP20ss fractions (MeOH–H<sub>2</sub>O, 50 : 50, and 90 : 10) from 550 mg extract of Gold Thread contain the 3 candidates with m/z of 320.08, 336.17, and 352.25. They were purified by semipreparative HPLC using a pentafluorophenyl column (isocratic Acetonitrile–H<sub>2</sub>O with 0.1% trifluoroacetic acid, 30 : 70), obtaining 3 pure compounds: berberine (**12**, 19.2 mg), palmatine (**13**, 5.4 mg), and coptisine (**14**, 4.9 mg). The <sup>1</sup>H and <sup>13</sup>C NMR data and spectra were shown in supporting information.

The fourth HP20ss fraction (MeOH–H<sub>2</sub>O, 90 : 10) from 1 g extract of *Stephania tetrandra* contains the 2 candidates with m/z of 609.33 and 623.33. They were purified by semipreparative HPLC using a C18 column (isocratic Acetonitrile–H<sub>2</sub>O with 0.02% diethylamine, 70 : 30), obtaining 2 pure compounds: fangchinoline (**15**, 1.2 mg), [ $\alpha$ ]<sub>D</sub> 275 (c 0.08, CHCl<sub>3</sub>) and tetrandrine (**16**, 7.2 mg), [ $\alpha$ ]<sub>D</sub> 279 (c 0.48, CHCl<sub>3</sub>). The <sup>1</sup>H NMR spectra were shown in supporting information.

The fourth HP20ss fraction (MeOH–H<sub>2</sub>O, 90 : 10) from 1 g extract of *Menispermum dauricum* contains the 2 candidates with m/z of 611.50 and 625.42. They were purified by semipreparative HPLC using a C18 column (isocratic Acetonitrile–H<sub>2</sub>O with 0.02% diethylamine,



60 : 40), obtaining 2 pure compounds: daurisolone (**17**, 1.2 mg),  $[\alpha]_D -125$  (c 0.08, MeOH) and dauricine (**18**, 2.3 mg),  $[\alpha]_D -134$  (c 0.15, MeOH). The  $^1\text{H}$  NMR spectra were shown in supporting information.

## **Chapter 4: Finding Biomass-degrading Enzymes Through an Activity-Correlated Quantitative Protein Profiling Platform (ACPP)**

*This chapter was adapted from a paper with the same title that was published on Journal of the American Society for Mass Spectrometry in 2017. The reuse licenses of this paper were issued by its copyright owners Springer Nature on Jun 20, 2019 (Supporting information, Appendix), and American Chemical Society on Jun 15, 2020 (Supporting information, Appendix). The copies of the reuse licenses have been submitted to Graduate College, and Bizzell Library of the University of Oklahoma as well.*

*The authors of this paper are Hongyan Ma, Daniel G. Delafield, Zhe Wang, Jianlan You, and Si Wu. The work presented within this chapter was conducted as follows: Daniel G. Delafield and Zhe Wang helped with part of preliminary work related to biomass-degrading enzyme activity measurement; Jianlan You helped with fungal culturing setup; as my previous advisor, Si Wu supervised all the work presented in this chapter. Hongyan Ma performed micro-scale biomass-degrading enzyme activity assay development, fungal secretome preparation, HPLC prefractionation, activity measurement, LC-MS-MS data collection, and quantitative proteomics data analysis.*

### **4.1 Introduction**

As an alternative energy source, bioethanol has attracted substantial attention due to its sustainability and renewability compared with traditional fuels. A key step of its production is converting plant biomass (e.g. starch and celluloses) into fermentable monomer sugars (e.g. glucose). Prior to publication of this research in 2017, this step mainly relies on extracellular fungal secretome (e.g. *Aspergillus*, *Trichoderma*, and *Fusarium*), which contains many different glycosyl

hydrolases (GHs), such as cellulases, breaking polysaccharides into monosaccharides. However, the commonly used fungal secretome cannot provide enough efficiency in industrial scale. On the other hand, thousands of proteins have been predicted with GH hydrolysis activities based on their genome sequences, many of which are potential highly-efficient biomass-degrading enzymes.<sup>192</sup> In addition, around 160,000 GH family sequences have been predicted by the CAZy database, in which only 6% of the predicted CAZymes were biochemically characterized till September 2013.<sup>193</sup> Thus, there is great potential to find new highly effective biomass-degrading enzymes in unknown fungal species.

Conventionally, functional protein (e.g. biomass-degrading enzyme) discovery method relies on targeted gene overexpression in host cells (e.g. *Escherichia coli* or yeast). Specifically, the overexpressed enzyme candidate is usually purified through several sequential liquid chromatography separations prior to further physical features characterization (e.g. size, sequence) and enzymatic activity analysis.<sup>82,194,195</sup> This approach can be readily incorporated into protein structure analysis and provide direct insights into activities of the target protein. Nevertheless, this method is limited by being relatively low throughput as it only allows one protein to be analyzed per screening. As to relatively high-yield enzymes, direct purification from crude fungal secretome is a common investigating method. For instance, with nine fast protein liquid chromatography (FPLC) separations followed by activity measurement on all collected fractions, Olsson's group successfully purified five cellulases from *Penicillium brasilianum* secretome against different hydrolysis substrates (microcrystalline cellulose, carboxymethyl cellulose, galactomannan, and xylan).<sup>196</sup> This method is typically labor-intensive and low throughput, and often requires large amount of starting materials. Recently, with emergence and development of LC-MS/MS-based proteomics, it is accessible to high-throughput identification of up to hundreds to thousands of proteins in

complex biological samples through several LC-MS runs. This approach has been applied to discover novel biomass-degrading enzymes in fungal secretome. For example, the Sze group has adopted an isobaric tag for absolute quantitation (iTRAQ)-based quantitative proteomics approach to analyze the expression levels of putative GH enzymes under different fungi culture conditions, providing key insights on enzyme mixture design for optimal biomass hydrolysis.<sup>197</sup> The enzyme candidates were predicted through comparing their expression levels with enzymatic activity of the whole secretome. This approach is lacking accuracy to assign enzymatic activity to individual protein because the expression level of each enzyme cannot be directly correlated to its own activity. Recently, the activity-based protein profiling (ABPP) approach has been applied to biomass-degrading enzymes profiling in fungal secretome.<sup>198,199,200</sup> Activity-based probes (ABPs) were used to covalently label and affinity-purify the GHs from fungal secretome for proteomics studies. The ABPP approach has been used to detect active biomass-degrading enzymes in different microbial organisms. Currently, this method is limited by the probe availability, in which, development, evaluation and optimization of GH-ABPs are usually cumbersome. Therefore, there is a critical need to develop a new high-throughput approach to directly assign the protein-level enzymatic activity.

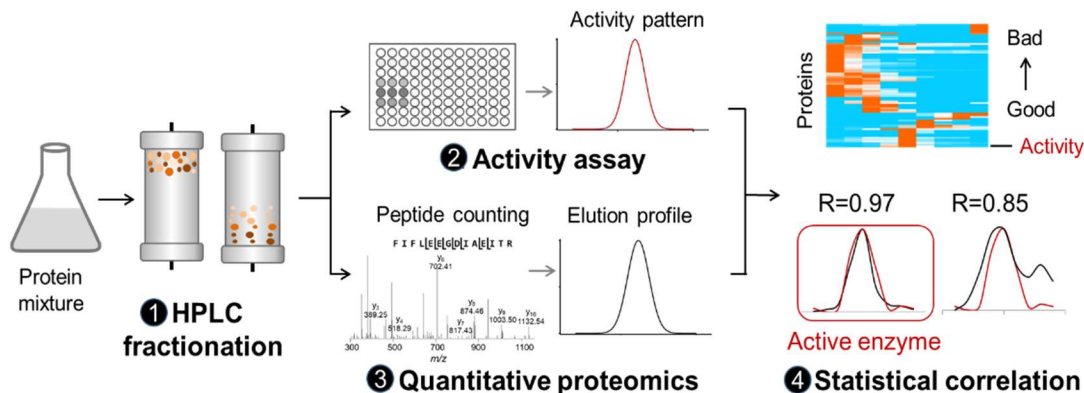
In this work, we developed an Activity-Correlated Quantitative Protein Profiling Platform (ACPP) that can systematically correlate protein-level enzymatic activity patterns with protein elution profiles obtained from the “native” LC separation and quantitative proteomics analysis to characterize bioactive enzymes from highly complex biological samples, in a high throughput and untargeted fashion. This platform consists of the 4 steps: 1) prefractionation in anion exchange chromatography (AEX), which provides good separation efficiency while preserving protein activity; 2) enzymatic activity pattern determination through a micro-scale (e.g. 5~100  $\mu$ L) activity

assay, which allows multiple enzyme activity characterization using one set of fractions; 3) protein elution profiles generation by LC-MS/MS-based quantitative proteomics, using peptide counting, a simple and robust label-free quantitative approach;<sup>201</sup> 4) statistical correlation between the detected enzymatic activity patterns and protein elution profiles, using Pearson Correlation algorithm. The correlation R score (-1 to 1) is used for enzyme candidate evaluation. The enzyme concentration is proportional to its activity in the linear dynamic range. Thus, the theoretical correlation between enzyme elution profile and its activity pattern should have a perfect R-score of 1. The ACPP was successfully applied to the characterization of biomass-degrading enzymes from the *Aspergillus niger* secretome, demonstrating this approach is a powerful tool to characterize new bioactive enzymes through the cross-correlation of activity patterns and protein elution patterns in an untargeted and high-throughput manner.

## **4.2 Results and Discussion**

### **4.2.1 ACPP Development and Proof-of-Concept**

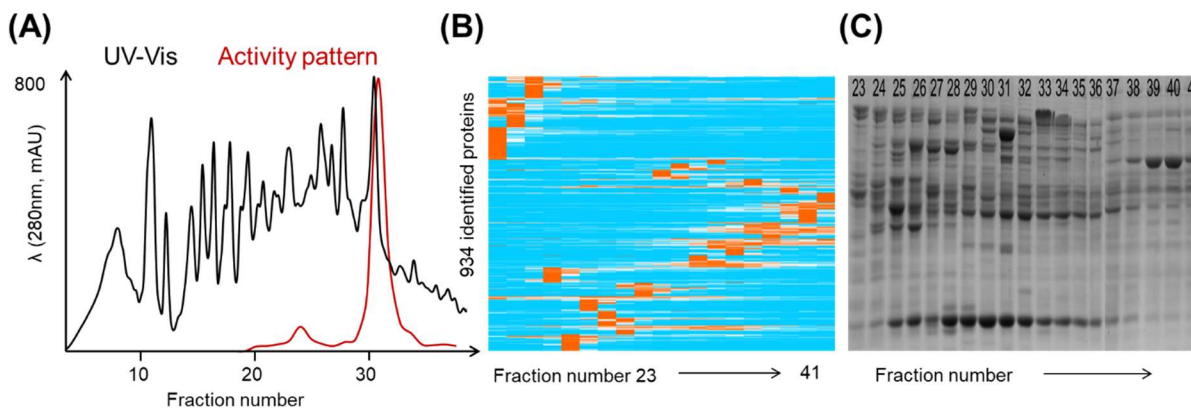
The overall ACPP strategy is demonstrated in Figure 4-1. In general, the ACPP method is composed of four steps: 1) “native” HPLC fractionation of the complex samples, which preserves enzymatic activities; 2) micro-scale bioactivity assay for enzymatic activity pattern generation; 3) quantitative protein profiling in which protein elution profiles are generated; and 4) statistical correlations between enzymatic activity patterns and the elution profiles of identified proteins. As a proof-of-principle, we examined the commercially available standard starch hydrolysis enzyme 1,4- $\alpha$ -glucosidase (AAG) from *A. niger* secretome. Performance and reproducibility for each step were first evaluated.



**Figure 4-1.** Overall workflow of the Activity-Correlated Quantitative Protein Profiling Platform (ACPP)

#### 4.2.1.1 HPLC Fractionation

The ACPP starts with a “native” LC fractionation that can preserve protein activities. Different chromatographic techniques have been applied in functional protein purification, including size exclusion chromatography (SEC) and ion exchange chromatography (IEX).<sup>202</sup> Here we adopted a high-performance anion exchange (AEX) LC using a GE (Pittsburgh, PA, USA) Mono Q column for the prefractionation of complex protein samples. We first evaluated the separation performance of complex protein samples (Figure 4-2) using SDS-PAGE and quantitative proteomics studies. The results revealed that the AEX separation technique using a GE MonoQ column can efficiently fractionate complex protein samples in ACPP. The following bioactivity assay across the collected fractions further confirmed that the selected “native” buffers in AEX can preserve enzymatic activities.

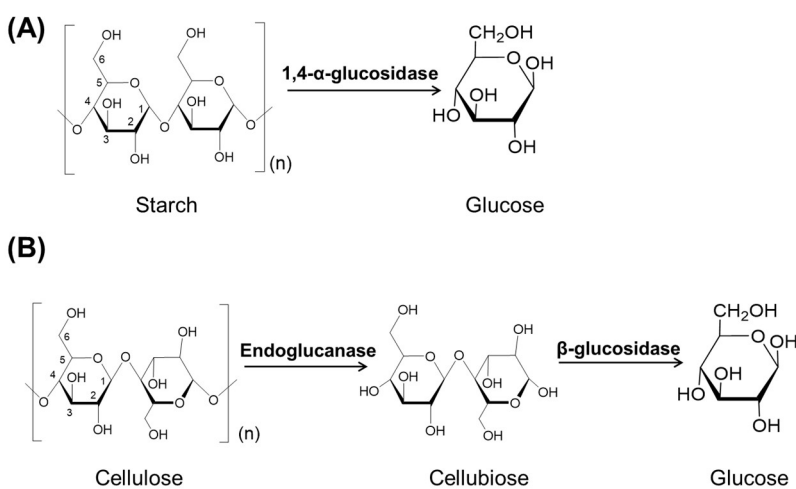


**Figure 4-2.** Analysis results of the *E.coli*. cell lysate spiked with standard AAG: A) the UV chromatogram and activity pattern from the mono Q based LC separation; B) hierarchical clustering of protein elution profiles on the basis of their similarities; C) SDS PAGE gel image of fractions 23 to 41

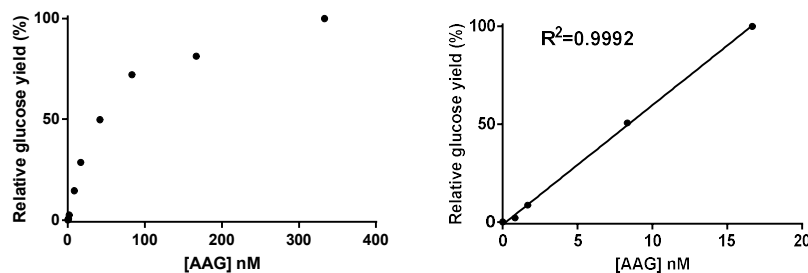
#### 4.2.1.2 Bioactivity Assay

Two biomass substrates were selected in this study: starch and cellulose. Starch is a linear polymer of glucose linked with (1→4) bonds, and cellulose is a polymer with repeated beta-linked glucose units (Figure 4-3).<sup>203</sup> Starch hydrolysis enzymes such as 1, 4- $\alpha$ -Glucosidase (AAG) convert starch into glucose, whereas cellulases catalyze the decomposition of cellulose into glucose monomers or smaller oligomers. Here we developed a two-step biomass-degrading enzyme activity assay: enzymatic hydrolysis reaction and glucose assay. First, an aliquot of the fractions (e.g. 5–100  $\mu$ L) was incubated with substrate solution under optimized pH and temperature conditions for optimal hydrolysis. Then, the supernatant of the reaction solution (~20  $\mu$ L) of each fraction was added to glucose fermentation enzyme mixture in a 96-well plate for high throughput micro-scale glucose detection. Here, the assay protocol was developed from a commercial glucose assay kit (Sigma-Aldrich), based on manufacturer’s instructions. In each analysis, a serial dilution of standard glucose solution was incubated with a premixed enzyme mixture to generate a standard calibration curve for the quantification of released glucose by each fraction. Quantification sensitivity,

reproducibility, and the linear dynamic range were first evaluated using a serial dilution of standard starch hydrolysis enzyme AAG (Figure 4-4). We further applied the developed activity assay on the fractions of the *E. coli* cell lysate spiked with standard AAG (e.g. 150  $\mu\text{g}$  AAG in 1.85 mg cell lysate). The assay was conducted using 5  $\mu\text{L}$  of each fraction (e.g. 1/200 of each fraction) twice at day 1 and day 30, storing the fractions at  $-80\text{ }^\circ\text{C}$  between these timepoints. The results demonstrated that the activity patterns of AAG showed very good reproducibility (Figure 4-5). Overall, the developed micro-scale activity assay is sensitive and highly reproducible, therefore can be used for generating accurate enzymatic activity patterns in ACPP.

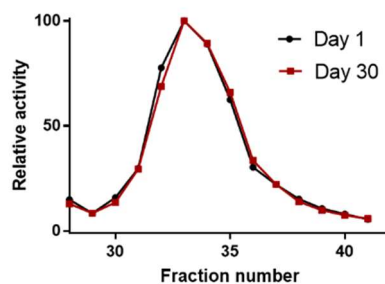


**Figure 4-3.** Hydrolysis mechanisms of starch and cellulose: A) 1, 4- $\alpha$ -Glucosidase (AAG) is a typical enzyme that can directly convert starch into glucose; B) cellulose is converted into glucose in a two-step reaction. Activity of the endo-cellulase (e.g. endoglucanase) can be measured in the presence of sufficient  $\beta$ -glucosidase.



**Figure 4-4.** The linear dynamic range analysis of the AAG activity using the developed two-step activity assay





**Figure 4-5.** Reproducibility performance of the activity assay towards the fractions of the *E. coli* cell lysate spiked with standard AAG (e.g. 150 µg AAG in 1.85 mg cell lysate).

#### 4.2.1.3 Proteome Profiling to Generate Protein Elution Profiles

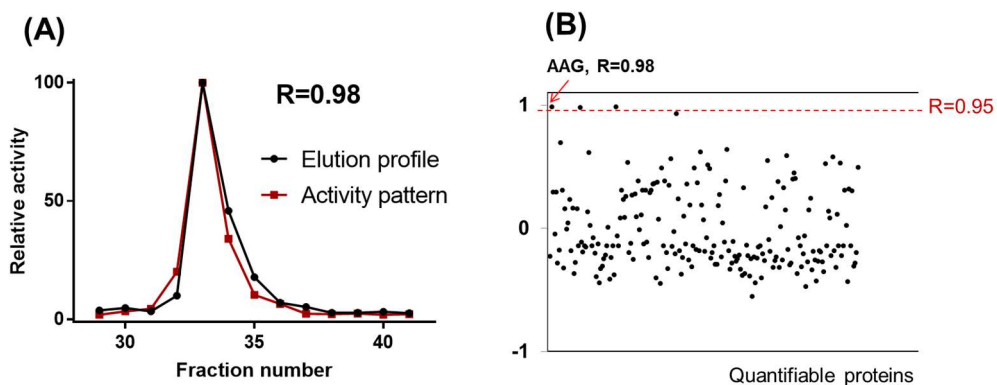
Protein identification and quantification in each fraction was conducted through untargeted LC-MS/MS analysis. Peptide counting, a label-free quantitation approach, was used to quantify the identified proteins to generate their elution profiles. As a simple and powerful quantitative proteomics technique, it has been applied to characterize commercial cellulase cocktails, demonstrating its capability for relative protein quantification among different samples.<sup>204</sup>

Label-free (e.g. peptide counting) quantitative proteomics analysis was performed on AEX fractions 23 to 41 of the *E. coli* lysate spiked with AAG. The results are illustrated in Figure 4-2. In total, 934 proteins were identified, and their elution profiles were generated by plotting the normalized protein intensities against their corresponding fraction numbers. The elution profiles can be visualized in a heat map format through hierarchical clustering using an open-source software. We then compared the protein elution profiles in SDS-PAGE analysis. Although AAG merged very well in the lysate on the gel, which can barely be observed, it was successfully identified and quantified by the adopted quantitative proteomics studies. In summary, the selected label-free quantitative proteome profiling approach can detect relatively low abundant proteins and provide reasonable elution profiles in ACPP.

#### 4.2.1.4 Statistical Correlation Between Activity Pattern and Protein Elution Profiles

In ACP, the bioactivity can be assigned through evaluating the statistical correlations between the activity pattern and protein elution profiles. The detected enzymatic activity pattern in sequential fractions was cross-correlated with all the identified protein elution profiles using Pearson Correlation algorithm, generating correlation R-scores range from +1 to -1. Theoretically, the enzyme concentrations are proportional to its activity in its linear dynamic range, so the correlation between enzyme elution profile and its activity pattern should have a perfect R-score of 1.

As a proof-of-concept, we first evaluated the correlation approach in *E.coli* lysate spiked with AAG sample. The elution profile of AAG correlates the best with the detected activity pattern ( $R = 0.98$ , Figure 4-6A). We then analyzed all the R-scores between the activity pattern and all the quantifiable protein elution profiles (e.g. total peptide counts greater than 10). As shown in Figure 4-6B, 3 proteins including AAG have best R-scores that are larger than 0.95. Their predicted activities were then manually checked through CAZY and NCBI databases, and only AAG is predicted with starch hydrolysis activity. To further confirm the results, we manually checked the top best-correlated proteins (e.g. the best 10 correlations). Among them, only AAG was confirmed as the starch hydrolysis enzyme through function verification by searching against the database of NCBI (<https://www.ncbi.nlm.nih.gov/protein/>).



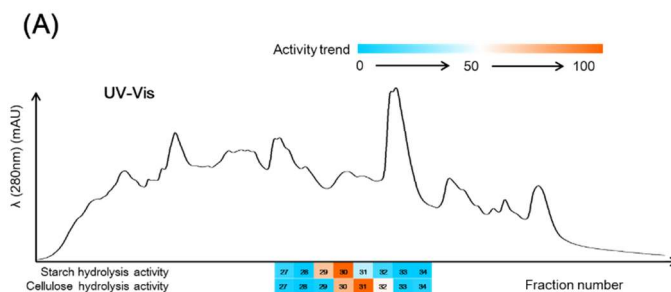
**Figure 4-6.** Proof-of-concept results: starch hydrolysis enzyme characterization from the *E. coli* cell lysate spiked with standard AAG. A) Overlap of the starch hydrolysis activity pattern and the AAG protein elution profile; B) Analysis of the correlations between the activity pattern and all quantifiable elution profiles.

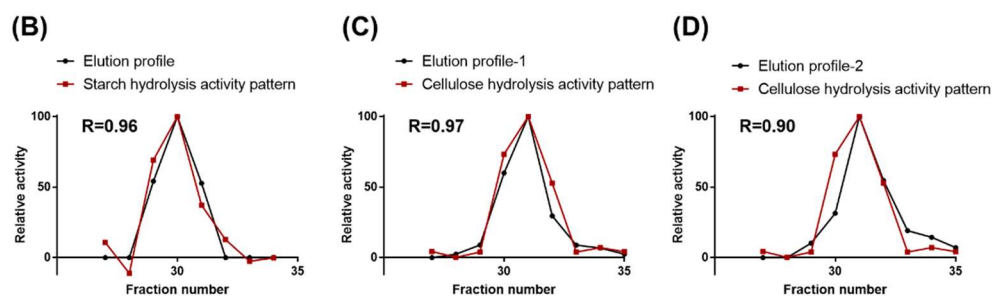
In summary, the enzymatic activity can be used to assign through the cross-correlation between the activity pattern and all identified protein elution profiles (e.g. R-scores). Nevertheless, two unrelated proteins give relatively high R-scores in ACPP as well. This observation can be attributed to the co-elution effect in LC separation, which is inevitable for complex samples in any one-dimensional chromatography. Thanks to the adopted high performance AEX separation column, the co-elution effect is relatively low, so ACPP can provide sufficient capability for enzyme annotation. Notably, manual verification of good-correlated proteins (e.g. top 10) is usually needed for confident active protein identification in ACPP.

#### 4.2.2 Biomass-degrading Enzymes Characterization from Fungal Secretome using ACPP

We then applied the well-developed ACPP approach to identify the biomass-hydrolysis enzymes from lab-cultured fungal secretome of *A. niger*, a model fungal strain that generates several different biomass-degrading enzymes. Two biomass substrates were selected in this study: starch and cellulose. The results demonstrated that ACPP can be utilized to identify both starch

and cellulose hydrolysis enzymes in a multiplexed fashion with one set of proteomics data. Activities of the two enzymes were measured (Figure 4-7A). The results showed they have different activity patterns under the developed LC fractionation conditions. In starch hydrolysis activity analysis, a single Gaussian-distributed peak was observed from fraction 27 to 34 with a peak center at fraction 30. All fractions in the activity range were then subjected to protein identification and quantification analysis. In total, 156 proteins were identified from these eight fractions, including two enzyme candidates with R-scores larger than 0.95. The best-correlated protein ( $R=0.96$ ) (Figure 4-7B), was identified as AAG (GH 15, EC 3.2.1.3), which is a classic starch hydrolysis enzyme in biofuel research field.<sup>205</sup> We further manually checked all the identified 25 GH proteins predicted by CAZY database. Four of them with R-scores higher than 0.80 were listed in Table 4-1, including GH15 (i.e., AAG), GH36, GH32, and GH18 proteins. The predicted substrate for the detected GH36 protein is glycolipids or glycoproteins, and the GH32 protein was reported as an inulinase. The GH18 protein was a hypothetical protein with unknown substrates. We further ran BLASTp of its sequence against the NCBI database.<sup>206</sup> The result did not show any sequences homologous with known starch degrading enzymes. In addition, the R-score of this GH18 protein is only 0.87, which indicates that it does not have good correlation with the starch hydrolysis activity pattern. Overall, the results demonstrated that ACPP is a powerful approach for bioactive enzymes identification from complex biological samples.





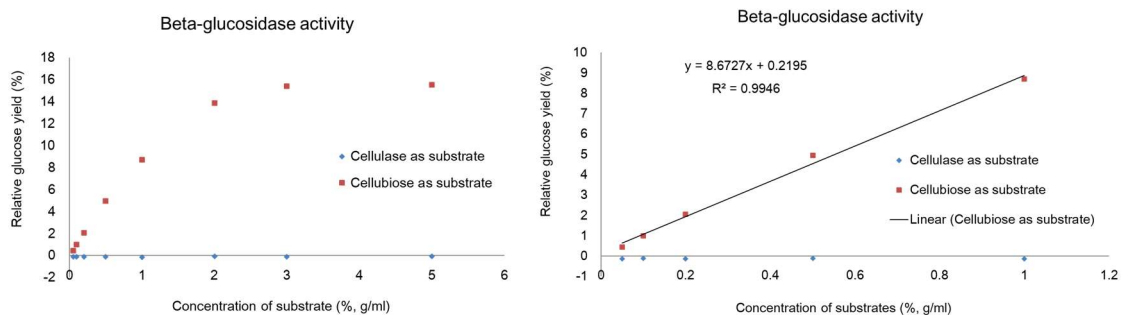
**Figure 4-7.** The results of applying ACPP into biomass-degrading enzymes identification: A) UV-Vis from the AEX fractionation (at 280nm); B) the correlation between the starch hydrolysis activity pattern and the best matched protein elution profile: AAG; and the correlations between the cellulose hydrolysis activity pattern and C) the best matched protein elution profile of endoglucanase and D) protein elution profile of  $\beta$ -glucosidase

**Table 4-1.** Manual check of the 4 starch hydrolysis enzyme candidates, which are belongs to glycosyl hydrolase (GH) family, with correlation R-scores higher than 0.8 in all 156 identified quantifiable proteins in the activity range (e.g. fraction 27-34)

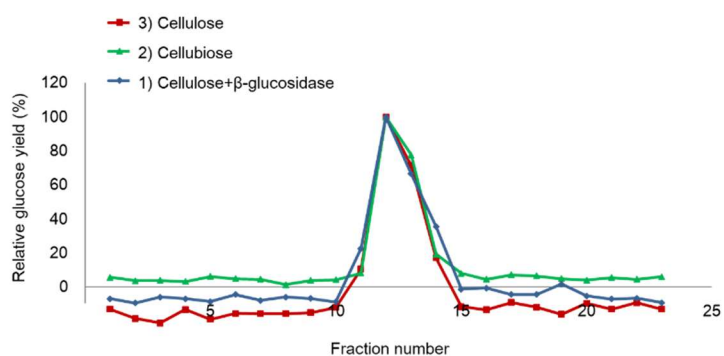
Protein ID	R score	GH family	Protein	Substrates
gi 350633017	0.96	15	AAG	starch
gi 350639761	0.95	36	$\alpha$ -galactosidase	Glycolipids or glycoproteins
gi 350631139	0.91	32	inulinase	inulin
gi 350640229	0.87	18	hypothetical protein	unknown

In cellulose hydrolysis enzyme characterization, the active range is the same as the starch hydrolysis activity, which is from 27 to 34 with a different peak center at fraction 31 (Figure 4-7c). In all detected proteins from the activity range, the best-correlated protein is endoglucanase (GH5) with R-score of 0.97. It is known that two steps involving two enzymes are necessary to

break cellulose to glucose monomers: endoglucanase and  $\beta$ -glucosidase (Figure 4-3). Endoglucanase can only convert cellulose to cellobiose, which is not detectable using the presented bioactivity assay.<sup>207</sup> The results indicated two possibilities: 1) the two enzymes were co-eluted under the adopted LC separation conditions, or 2) it represented a novel cellulase that can directly hydrolyze cellulose to glucose. With further investigation, the protein with second highest R-score of 0.90 is identified as  $\beta$ -glucosidase, which can convert cellobiose to glucose. Thanks to the presence of the co-eluted  $\beta$ -glucosidase, the endoglucanase activity was successfully detected via the presented activity assay. We further confirmed that  $\beta$ -glucosidase cannot convert cellulose into glucose using the commercial enzyme (Figure 4-8). Therefore, endoglucanase and  $\beta$ -glucosidase are the cellulose hydrolysis enzymes identified from *A. niger* secretome using the ACPP. To further confirm that the two proteins were co-eluted under the adopted LC conditions, we performed three sets of activity assays on three separated aliquots of HPLC fractions (e.g. 5-100  $\mu$ L) incubated with different substrates: 1) cellulose with the presence of over-dose commercial  $\beta$ -glucosidase, 2) cellobiose, and 3) cellulose. Theoretically, the first set provides elution profile of endoglucanase, the second set provides elution profile of  $\beta$ -glucosidase, and the third set evaluates if endoglucanase and  $\beta$ -glucosidase were co-eluted. The hypothesis is that if they are not co-eluted, the activity peak center should differentiate in set 1 and 2. As it shows in Figure 4-9, the 3 activity patterns overlapped very well with each other. The result demonstrated that the commercial cellulase cocktail contains both endoglucanase and  $\beta$ -glucosidase, which were co-eluted under the adopted AEX fractionation conditions. Besides, the co-eluted two enzymes were successfully identified as the cellulose hydrolysis enzymes using ACPP. In conclusion, the ACPP approach showed its power and robustness to identify bioactive proteins (e.g. enzymes) from complex biological samples in an untargeted, high throughput and multiplexed fashion.



**Figure 4-8.** Cellulose and cellubiose hydrolysis by commercial  $\beta$ -glucosidase.



**Figure 4-9.** Co-elution effect evaluation results. Three hydrolysis reactions were conducted: 1) cellulose + HPLC fractions + commercial  $\beta$ -glucosidase, 2) cellobiose + HPLC fractions, and 3) cellulose + HPLC fractions. The blue line with diamond represents the endoglucanase elution profile, the green line with triangle represents the  $\beta$ -glucosidase elution profile, and the red line with square evaluates if endoglucanase and  $\beta$ -glucosidase were co-eluted under the adopted AEX fractionation conditions.

### 4.2.3 Conclusions

In conclusion, ACPP is a powerful approach for identifying bioactive enzymes from complex protein samples by cross correlating the activity pattern and all identified protein elution profiles in the activity range (e.g. active fractions), in a high throughput and untargeted manner. It can identify more than one active enzymes from complicated protein mixtures, thanks to the micro-

scale activity assay and the sensitive quantitative proteomics technique. In addition, the ACP approach can be further optimized into a universal functional protein characterization approach for different applications, such as disease-regulating enzyme screening or drug target discovery.

## **4.3 Materials and Methods**

### **4.3.1 *A. niger* Fermentation and Secretome Extraction**

All chemicals were purchased from Sigma-Aldrich (Milwaukee, WI, USA) unless otherwise noted. *A. niger* (ATCC11414) was recovered on a PDA plate for further liquid fermentation as described previously.<sup>208</sup> After shaking at 37 °C, 200 rpm for 24 hours, the supernatant was harvested by filtering the cultured medium mixture through two layers of sterile miracloth, followed by centrifugation at 15,000 g and 4 °C for 10 min to remove the remaining cell debris. The harvested crude secretome was further concentrated to 20 mL in an ultrafiltration stirring cell at 4 °C (Millipore, Billerica, MA, USA) with a 3 kDa membrane. Buffer exchange was then conducted twice using Milli-Q water (e.g. from 250 to 20 mL) in the same stirring cell. The remaining secretome was further concentrated down to 3~5 mL using a 3 kDa centrifugal ultrafiltration filter (Millipore) by centrifugation. The prepared secretome was stored at -20 °C until being analyzed.

### **4.3.2 Anion Exchange LC Fractionation**

The LC fractionation was performed at 4 °C using an AKTA protein purification system (GE, Pittsburgh, PA, USA) with a high-performance mono-Q column (4.6 × 100 mm; GE). The secretome was exchanged to buffer A (20 mM Tris, pH = 8.0) before sample injection. Elution gradient was set as 0–100% buffer B (20 mM Tris, 1 M NaCl, pH = 8.0) in 38 column volume at



a flow rate of 0.5 mL/min. Fractions (0.83 mL per fraction for the spiked-in sample, 1 mL per fraction for the *A. niger* secretome, and 1.66 mL per fraction for *A. niger* cellulose and cellubiose activity assays) were collected for activity assays and proteomics studies. Three aliquots from each fraction were used for: (1) the biomass hydrolysis activity assay (e.g. 5–100  $\mu$ L), (2) the cellulose hydrolysis activity assay (e.g. 20–100  $\mu$ L), and (3) tryptic digestion and LC/MS/MS based bottom-up proteomic characterization (e.g. 100–200  $\mu$ L). The remaining fractions were stored at  $-20\text{ }^{\circ}\text{C}$  until further analysis.

### 4.3.3 Microplate-Based Enzymatic Activity Assays

A two-step 96-well plate-based enzymatic assay (e.g. substrate hydrolysis and glucose detection) was developed based on the reported large-scale (e.g. 4 mL) assay method<sup>209,210</sup> to achieve higher throughput and lower sample consumption. First, in substrate hydrolysis step, a small aliquot of each fraction (5–100  $\mu$ L) was mixed with 5% (w/v) cellulose, cellubiose, or starch substrates at the ratio of 1:4. The mixture was incubated at  $37\text{ }^{\circ}\text{C}$  and 900 rpm for 2 hours, followed by being quenched on ice for 10 min. Reaction supernatant was collected after centrifugation at 13000 g,  $4\text{ }^{\circ}\text{C}$ , for 30 min. The concentration of released glucose in each reaction supernatant was measured based on a commercial glucose assay protocol (Sigma-Aldrich). Briefly, the supernatant of reaction mixtures (e.g. 20  $\mu$ L) and glucose working reagent were first mixed at the ratio of 1:5 in a 96-well plate. After incubation at  $25\text{ }^{\circ}\text{C}$  for 30 min, the absorbance of the reaction mixture was measured at 340 nm. The activities of fractions were represented by the detected glucose concentration.

#### 4.3.4 LC-MS/MS Analysis

In-solution tryptic digestion was performed to a set of fraction aliquots (e.g. 100–200  $\mu\text{L}$ ) using a previously published protocol.<sup>211</sup> Peptide samples were desalted, dried via SpeedVac, and resuspended in ultra-pure water; 1/5 of each peptide sample was analyzed using a custom-packed C18 RPLC column (75  $\mu\text{m}$  i.d., 150 mm length, 2  $\mu\text{m}$  C18 resin) on a Waters (Milford, MA, USA) nano-Acquity UPLC system. The peptide separation was performed in a water-acetonitrile LC system, with buffer A of 0.1% formic acid in water, and buffer B of 0.1% formic acid in acetonitrile. The gradient was set from 8% to 35% buffer B over 100 mins. The coupled mass spectrometer was an LTQ Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific, Hanover Park, IL, USA) with a custom nano-ESI interface. The capillary temperature was set to 300  $^{\circ}\text{C}$  with a spray voltage of 2.4 kV. Full MS spectra were acquired at a resolution of 30000 ( $m/z$  range between 350 and 2000). Low resolution collisional induced dissociation (CID) MS/MS with a normalized collision energy setting of 35% was applied in data dependent acquisition (DDA) mode to analyze the most ten abundant parent ions.

#### 4.3.5 Peptide and Protein Identification

Peptides and proteins were identified using an online platform MSGF+ to search the mass spectra from LC-MS/MS analysis against the annotated *A. niger* database and its decoy database.<sup>212</sup> Peptide identifications were filtered using an MS-GF cut-off value of  $1\text{E}-10$  (e.g. the calculated FDR <1% at the unique peptide level). Spectral counting, namely peptide counting approach was used for label-free quantitative proteomics analysis. The normalized quantitation of the same protein in different fractions was plotted against the fraction number to obtain the protein elution profile.

#### 4.3.6 Statistical Correlation between Protein Profiles and Activity Pattern

Cross-correlation between protein elution profiles and the activity pattern was obtained through a linear correlation algorithm—Pearson Correlation. The correlation efficiency was represented by the calculated R-score, which is from  $-1$  to  $1$ . The protein elution profiles were refined using a peptide count number cut-off of 10 to distinguish a good elution profile from random noises. Before the cross-correlation calculation, individual values in each protein elution profile and activity pattern were normalized respectively. Thus, an R of 1 indicates the protein elution profile is perfectly overlapped with the activity pattern. The top correlated proteins with high R values were manually verified with their estimated activity to confidently characterize the enzyme from complicated protein mixture.

#### 4.4 Future directions of the ACPP

*Here, further improvements of ACPP, such as 2D-ACPP and TMT-ACPP approaches, and their successful applications were briefly summarized. This work has led to a Lloyd E. Swearingen Scholarship for the recognition of the outstanding graduate research in the Department of Chemistry & Biochemistry, and 3 travel grants to international academic conferences from the University of Oklahoma. Moreover, 1 oral and 4 poster presentations about the 2D-ACPP and TMT-ACPP work have been presented at different international conferences, including ASMS, USHUPO, and ISCC. Unfortunately, due to the conflict of interest, the data cannot be presented in detail here.*

We envision potential possibilities to enhance the resolution and accuracy of current **ACPP** approach. *First*, as previously mentioned, the co-elution effect is inevitable in any one-dimensional

chromatography fractionation, leading to the observation that several unrelated inactive proteins stand out in the correlation step. One possible solution is to add an additional dimension to the fractionation (e.g. size exclusion) to remove the enzyme candidate noise. The additional dimension can also facilitate the expansion of proteome coverage, by identifying more low abundance proteins. Based on this concept, I further developed 2D-ACPP to expand the ACPP capability by increasing the resolution for functional protein characterization through incorporating an additional orthogonal fractionation [e.g. size exclusive chromatography (SEC)] step. Through the successful characterization of both starch and cellobiose hydrolysis enzymes in fungal secretome spiked with *E.coli* lysate with no random hit, it is believed that the **2D-ACPP** approach can effectively remove the randomly matched protein candidates for functional enzyme characterization in highly complex biological samples, due to the good orthogonality between SEC and AEX fractionation techniques.

*Second*, the protein elution profiles were generated using peptide counting-based label-free quantitative proteomics technique in ACPP. Peptide counting is a simple and robust technique for relative protein quantification among samples yet is often restricted by the linear dynamic range for low abundance protein quantification. Isotope labeling quantitative methods, such as tandem mass tag (TMT), provides higher accuracy for protein quantification, especially for the low abundance proteins in complex protein samples. The TMT labeling quantitative proteomics approach can directly correlate the reporter ion profiles with enzyme activity patterns for deeper identification and quantification of low abundance proteins. Besides, the peptide-level 2D separation approach (e.g. high-pH-RPLC coupling with low-pH RPLC) has been proven that it can largely enhance the proteome coverage in proteomics studies. Therefore, the **TMT-ACPP** coupling with the

peptide-level 2D separation technique was developed for generation of more accurate protein elution profiles, especially for those proteins with low abundance. The TMT-ACPP has been successfully developed and applied for functional characterization of active kinases (e.g. important cancer drug targets) from highly complex human cancer cell lysates (e.g. HeLa cell line).

It is envisioned that through further developments, the ACPP will enable efficient characterization of functional proteins in various biological and bioanalytical fields.

#### **Acronym list of chapter 4:**

1, 4- $\alpha$ -Glucosidase	AAG
activity-based protein profiling	ABPP
activity-based probes	ABPs
Activity-Related Quantitative Protein Profiling Platform	ACPP
anion exchange chromatography	AEX
fast protein liquid chromatography	FPLC
glycosyl hydrolases	GHs
ion exchange chromatography	IEX
isobaric tag for absolute quantitation	iTRAQ
size exclusion chromatography	SEC
tandem mass tag	TMT

## References

1. Sheikh, T.; Shinde, A., A review and recent update on LCMS.
2. Molnar, I.; Horvath, C., Reverse-phase chromatography of polar biological substances: separation of catechol compounds by high-performance liquid chromatography. *Clin Chem* **1976**, *22* (9), 1497-502.
3. MacNair, J. E.; Lewis, K. C.; Jorgenson, J. W., Ultrahigh-pressure reversed-phase liquid chromatography in packed capillary columns. *Anal Chem* **1997**, *69* (6), 983-9.
4. Wu, N.; Lippert, J. A.; Lee, M. L., Practical aspects of ultrahigh pressure capillary liquid chromatography. *J Chromatogr A* **2001**, *911* (1), 1-12.
5. Dong, M. W., Ultrahigh-pressure LC in pharmaceutical analysis: performance and practical issues. *LC GC North America* **2007**, 89-98.
6. Robards, K.; Robards, K.; Haddad, P. R.; Haddad, P. R.; Jackson, P.; Jackson, P.; Haddad, P. A., *Principles and practice of modern chromatographic methods*. Academic Press: 1994.
7. Sparkman, O. D., Mass spectrometry desk reference. **2006**.
8. Vehovec, T.; Obreza, A., Review of operating principle and applications of the charged aerosol detector. *J Chromatogr A* **2010**, *1217* (10), 1549-56.
9. Douville, V.; Lodi, A.; Miller, J.; Nicolas, A.; Clarot, I.; Prilleux, B.; Megoulas, N.; Koupparis, M., Evaporative light scattering detection (ELSD): a tool for improved quality control of drug substances. *Pharmeur Sci Notes* **2006**, *2006* (1), 9-15.
10. Pitt, J. J., Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev* **2009**, *30* (1), 19-34.
11. Downard, K., *Mass spectrometry: a foundation course*. Royal Society of Chemistry: 2007.
12. Cottrell, J. S.; Greathead, R. J., Extending the mass range of a sector mass spectrometer. *Mass Spectrometry Reviews* **1986**, *5* (3), 215-247.
13. Pareige, C.; Lefebvre-Ulrikson, W.; Vurpillot, F.; Sauvage, X., Time-of-Flight mass spectrometry and composition measurements. In *Atom Probe Tomography*, Elsevier: 2016; pp 123-154.
14. Douglas, D., Linear quadrupoles in mass spectrometry. *Mass spectrometry reviews* **2009**, *28* (6), 937-960.
15. Nolting, D.; Malek, R.; Makarov, A., Ion traps in modern mass spectrometry. *Mass Spectrom Rev* **2019**, *38* (2), 150-168.
16. Marshall, A. G.; Grosshans, P. B., Fourier transform ion cyclotron resonance mass spectrometry: the teenage years. *Analytical Chemistry* **1991**, *63* (4), 215A-229A.
17. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* **1998**, *17* (1), 1-35.
18. MR, L. H.; Gopinath, C., A review on GC-MS and method development and validation. *International Journal of Pharmaceutical Quality Assurance* **2013**, *4* (03), 42-51.
19. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246* (4926), 64-71.
20. Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R., New developments in biochemical mass spectrometry: electrospray ionization. *Analytical Chemistry* **1990**, *62* (9), 882-899.
21. Mazloumian, A.; Eom, Y.-H.; Helbing, D.; Lozano, S.; Fortunato, S., How citation boosts promote scientific paradigm shifts and nobel prizes. *Plos One* **2011**, *6* (5).
22. Wilm, M., Principles of electrospray ionization. *Mol Cell Proteomics* **2011**, *10* (7).
23. Banerjee, S.; Mazumdar, S., Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International journal of analytical chemistry* **2012**, *2012*.
24. Ho, C. S.; Lam, C. W.; Chan, M. H.; Cheung, R. C.; Law, L. K.; Lit, L. C.; Ng, K. F.; Suen, M. W.; Tai, H. L., Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev* **2003**, *24* (1), 3-12.
25. Billat, P. A.; Saint-Marcoux, F., Liquid chromatography-mass spectrometry methods for the intracellular determination of drugs and their metabolites: a focus on antiviral drugs. *Anal Bioanal Chem* **2017**, *409* (25), 5837-5853.
26. van Eijk, H. M.; Bloemen, J. G.; Dejong, C. H., Application of liquid chromatography-mass spectrometry to measure short chain fatty acids in blood. *Journal of Chromatography B* **2009**, *877* (8-9), 719-724.
27. Zwiener, C.; Frimmel, F. H., LC-MS analysis in the aquatic environment and in water treatment technology-a critical review. Part II: Applications for emerging contaminants and related pollutants, microorganisms and humic

acids. *Anal Bioanal Chem* **2004**, 378 (4), 862-74.

28. Hernandez, F.; Sancho, J.; Pozo, O., Critical review of the application of liquid chromatography/mass spectrometry to the determination of pesticide residues in biological samples. *Analytical and bioanalytical chemistry* **2005**, 382 (4), 934-946.
29. Nicholson, J. K.; Lindon, J. C.; Holmes, E., 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999**, 29 (11), 1181-9.
30. Johnson, C. H.; Gonzalez, F. J., Challenges and opportunities of metabolomics. *J Cell Physiol* **2012**, 227 (8), 2975-81.
31. Johnson, C. H.; Patterson, A. D.; Idle, J. R.; Gonzalez, F. J., Xenobiotic metabolomics: major impact on the metabolome. *Annu Rev Pharmacol Toxicol* **2012**, 52, 37-56.
32. Emwas, A. H.; Roy, R.; McKay, R. T.; Tenori, L.; Saccenti, E.; Gowda, G. A. N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; Wishart, D. S., NMR spectroscopy for metabolomics research. *Metabolites* **2019**, 9 (7).
33. Tsugawa, H.; Tsujimoto, Y.; Sugitate, K.; Sakui, N.; Nishiumi, S.; Bamba, T.; Fukusaki, E., Highly sensitive and selective analysis of widely targeted metabolomics using gas chromatography/triple-quadrupole mass spectrometry. *J Biosci Bioeng* **2014**, 117 (1), 122-8.
34. Zhao, Y.; Hao, Z.; Zhao, C.; Zhao, J.; Zhang, J.; Li, Y.; Li, L.; Huang, X.; Lin, X.; Zeng, Z.; Lu, X.; Xu, G., A novel strategy for large-scale metabolomics study by calibrating gross and systematic errors in gas chromatography-mass spectrometry. *Anal Chem* **2016**, 88 (4), 2234-42.
35. Soga, T., Capillary electrophoresis-mass spectrometry for metabolomics. *Methods Mol Biol* **2007**, 358, 129-37.
36. Rodrigues, K. T.; Mekahli, D.; Tavares, M. F.; Van Schepdael, A., Development and validation of a CE-MS method for the targeted assessment of amino acids in urine. *Electrophoresis* **2016**, 37 (7-8), 1039-47.
37. Shulaev, V.; Isaac, G., Supercritical fluid chromatography coupled to mass spectrometry-A metabolomics perspective. *J Chromatogr B Analyt Technol Biomed Life Sci* **2018**, 1092, 499-505.
38. Fang, Z. Z.; Gonzalez, F. J., LC-MS-based metabolomics: an update. *Arch Toxicol* **2014**, 88 (8), 1491-502.
39. Canelas, A. B.; ten Pierick, A.; Ras, C.; Seifar, R. M.; van Dam, J. C.; van Gulik, W. M.; Heijnen, J. J., Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. *Anal Chem* **2009**, 81 (17), 7379-89.
40. Wawrzyniak, R.; Kosnowska, A.; Macioszek, S.; Bartoszewski, R.; Markuszewski, M. J., New plasma preparation approach to enrich metabolome coverage in untargeted metabolomics: Plasma protein bound hydrophobic metabolite release with proteinase K. *Scientific reports* **2018**, 8 (1), 1-10.
41. Katajamaa, M.; Orešič, M., Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics* **2005**, 6 (1), 179.
42. Smith, C. A.; Want E J Fau - O'Maille, G.; O'Maille G Fau-Abagyan, R.; Abagyan R Fau-Siuzdak, G.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. (0003-2700 (Print)).
43. Lommen, A., MetAlign: Interface-Driven, Versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry* **2009**, 81 (8), 3079-3086.
44. Misra, B. B.; van der Hoof, J. J., Updates in metabolomics tools and resources: 2014-2015. *Electrophoresis* **2016**, 37 (1), 86-110.
45. Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L., HMDB: the human metabolome database. *Nucleic Acids Res* **2007**, 35 (Database issue), D521-6.
46. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T., MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* **2010**, 45 (7), 703-14.
47. Domingo-Almenara, X.; Montenegro-Burke, J. R.; Ivanisevic, J.; Thomas, A.; Sidibe, J.; Teav, T.; Guijas, C.; Aisporna, A. E.; Rinehart, D.; Hoang, L.; Nordstrom, A.; Gomez-Romero, M.; Whiley, L.; Lewis, M. R.; Nicholson, J. K.; Benton, H. P.; Siuzdak, G., XCMS-MRM and METLIN-MRM: a cloud library and public resource for targeted analysis of small molecules. *Nat Methods* **2018**, 15 (9), 681-684.

48. Chen, J.; Swamidass, S. J.; Dou, Y.; Bruand, J.; Baldi, P., ChemDB: a public database of small molecules and related cheminformatics resources. *Bioinformatics* **2005**, *21* (22), 4133-9.
49. Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W., LC-MS-based metabolomics. *Molecular BioSystems* **2012**, *8* (2), 470-481.
50. Yuan, M.; Breikopf, S. B.; Yang, X.; Asara, J. M., A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nat Protoc* **2012**, *7* (5), 872-81.
51. Zhou, J.; Liu, H.; Liu, Y.; Liu, J.; Zhao, X.; Yin, Y., Development and evaluation of a parallel reaction monitoring strategy for large-scale targeted metabolomics quantification. *Anal Chem* **2016**, *88* (8), 4478-86.
52. Lee, K. M.; Jeon, J. Y.; Lee, B. J.; Lee, H.; Choi, H. K., Application of metabolomics to quality control of natural product derived medicines. *Biomol Ther (Seoul)* **2017**, *25* (6), 559-568.
53. Demarque, D. P.; Dusi, R. G.; de Sousa, F. D. M.; Grossi, S. M.; Silverio, M. R. S.; Lopes, N. P.; Espindola, L. S., Mass spectrometry-based metabolomics approach in the isolation of bioactive natural products. *Sci Rep* **2020**, *10* (1), 1051.
54. Abuawad, A.; Mbadugha, C.; Ghaemmaghami, A. M.; Kim, D.-H., Metabolic characterisation of THP-1 macrophage polarisation using LC-MS-based metabolite profiling. *Metabolomics* **2020**, *16* (3), 1-14.
55. Simic, P.; Kim, W.; Zhou, W.; Pierce, K. A.; Chang, W.; Sykes, D. B.; Aziz, N. B.; Elmariah, S.; Ngo, D.; Pajevic, P. D.; Govea, N.; Kestenbaum, B. R.; de Boer, I. H.; Cheng, Z.; Christov, M.; Chun, J.; Leaf, D. E.; Waikar, S. S.; Tager, A. M.; Gerszten, R. E.; Thadhani, R. I.; Clish, C. B.; Juppner, H.; Wein, M. N.; Rhee, E. P., Glycerol-3-phosphate is an FGF23 regulator derived from the injured kidney. *J Clin Invest* **2020**, *130* (3), 1513-1526.
56. Shao, Y.; Zhu, B.; Zheng, R.; Zhao, X.; Yin, P.; Lu, X.; Jiao, B.; Xu, G.; Yao, Z., Development of urinary pseudotargeted LC-MS-based metabolomics method and its application in hepatocellular carcinoma biomarker discovery. *J Proteome Res* **2015**, *14* (2), 906-16.
57. Martín-Blázquez, A.; Díaz, C.; González-Flores, E.; Franco-Rivas, D.; Jiménez-Luna, C.; Melguizo, C.; Prados, J.; Genilloud, O.; Vicente, F.; Caba, O., Untargeted LC-HRMS-based metabolomics to identify novel biomarkers of metastatic colorectal cancer. *Scientific Reports* **2019**, *9* (1), 1-9.
58. Li, R.; Li, F.; Feng, Q.; Liu, Z.; Jie, Z.; Wen, B.; Xu, X.; Zhong, S.; Li, G.; He, K., An LC-MS based untargeted metabolomics study identified novel biomarkers for coronary heart disease. *Molecular BioSystems* **2016**, *12* (11), 3425-3434.
59. Milburn, M. V.; Lawton, K. A., Application of metabolomics to diagnosis of insulin resistance. *Annu Rev Med* **2013**, *64*, 291-305.
60. Dias, D. A.; Urban, S.; Roessner, U., A historical overview of natural products in drug discovery. *Metabolites* **2012**, *2* (2), 303-36.
61. Boufridi, A.; Quinn, R. J., Harnessing the properties of natural products. *Annu Rev Pharmacol* **2018**, *58*, 451-470.
62. Wolfender, J. L.; Nuzillard, J. M.; van der Hoof, J. J. J.; Renault, J. H.; Bertrand, S., Accelerating metabolite identification in natural product research: toward an ideal combination of liquid chromatography-high-resolution tandem mass spectrometry and NMR profiling, in silico databases, and chemometrics. *Anal Chem* **2019**, *91* (1), 704-742.
63. Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G., XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem* **2012**, *84* (11), 5035-9.
64. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010**, *11*, 395.
65. Rost, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H. C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.; Choudhary, J. S.; Malmstrom, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O., OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods* **2016**, *13* (9), 741-8.
66. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* **2015**, *12* (6), 523-6.
67. Clasquin, M. F.; Melamud, E.; Rabinowitz, J. D., LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr Protoc Bioinformatics* **2012**, *Chapter 14*, Unit14 11.
68. Giacomoni, F.; Le Corguille, G.; Monsoor, M.; Landi, M.; Pericard, P.; Petera, M.; Duperier, C.; Tremblay-Franco, M.; Martin, J. F.; Jacob, D.; Goullitquer, S.; Thevenot, E. A.; Caron, C., Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics* **2015**, *31* (9), 1493-5.



69. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J., MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* **2018**, *46* (W1), W486-W494.
70. Quinn, R. A.; Nothias, L.-F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P. C., Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy. *Trends Pharmacol Sci* **2017**, *38* (2), 143-154.
71. Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L. F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J. M.; Sikora, N.; Vargas, F.; van der Hoof, J. J. J.; Ernst, M.; Kang, K. B.; Aceves, C. M.; Caraballo-Rodriguez, A. M.; Koester, I.; Weldon, K. C.; Bertrand, S.; Roullier, C.; Sun, K.; Tehan, R. M.; Boya, P. C.; Christian, M. H.; Gutierrez, M.; Ulloa, A. M.; Tejada Mora, J. A.; Mojica-Flores, R.; Lakey-Beitia, J.; Vasquez-Chaves, V.; Zhang, Y.; Calderon, A. I.; Tayler, N.; Keyzers, R. A.; Tugizimana, F.; Ndlovu, N.; Aksenov, A. A.; Jarmusch, A. K.; Schmid, R.; Truman, A. W.; Bandeira, N.; Wang, M.; Dorrestein, P. C., Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat Protoc* **2020**, *15* (6), 1954-1991.
72. Zani, C. L.; Carroll, A. R., Database for rapid dereplication of known natural products using data from MS and fast NMR experiments. *J Nat Prod* **2017**, *80* (6), 1758-1766.
73. van Santen, J. A.; Jacob, G.; Singh, A. L.; Aniebok, V.; Balunas, M. J.; Bunsko, D.; Neto, F. C.; Castano-Espriu, L.; Chang, C.; Clark, T. N.; Cleary Little, J. L.; Delgadillo, D. A.; Dorrestein, P. C.; Duncan, K. R.; Egan, J. M.; Galey, M. M.; Haeckl, F. P. J.; Hua, A.; Hughes, A. H.; Iskakova, D.; Khadilkar, A.; Lee, J. H.; Lee, S.; LeGrow, N.; Liu, D. Y.; Macho, J. M.; McCaughey, C. S.; Medema, M. H.; Neupane, R. P.; O'Donnell, T. J.; Paula, J. S.; Sanchez, L. M.; Shaikh, A. F.; Soldatou, S.; Terlouw, B. R.; Tran, T. A.; Valentine, M.; van der Hoof, J. J. J.; Vo, D. A.; Wang, M.; Wilson, D.; Zink, K. E.; Linington, R. G., The natural products atlas: an open access knowledge base for microbial natural products discovery. *ACS Cent Sci* **2019**, *5* (11), 1824-1833.
74. Tawfike, A. F.; Viegelmann, C.; Edrada-Ebel, R., Metabolomics and dereplication strategies in natural products. *Methods Mol Biol* **2013**, *1055*, 227-44.
75. Liaw, C. C.; Chen, P. C.; Shih, C. J.; Tseng, S. P.; Lai, Y. M.; Hsu, C. H.; Dorrestein, P. C.; Yang, Y. L., Vitroprocines, new antibiotics against *Acinetobacter baumannii*, discovered from marine *Vibrio* sp. QWI-06 using mass-spectrometry-based metabolomics approach. *Sci Rep* **2015**, *5*, 12856.
76. Bertin, M. J.; Schwartz, S. L.; Lee, J.; Korobeynikov, A.; Dorrestein, P. C.; Gerwick, L.; Gerwick, W. H., Spongiosine production by a *Vibrio harveyi* strain associated with the sponge *Tectitethya crypta*. *J Nat Prod* **2015**, *78* (3), 493-9.
77. Yu, H. B.; Glukhov, E.; Li, Y.; Iwasaki, A.; Gerwick, L.; Dorrestein, P. C.; Jiao, B. H.; Gerwick, W. H., Cytotoxic microcolin lipopeptides from the marine cyanobacterium *Moorea producens*. *J Nat Prod* **2019**, *82* (9), 2608-2619.
78. 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.
79. Demarque, D. P.; Dusi, R. G.; de Sousa, F. D.; Grossi, S. M.; Silvério, M. R.; Lopes, N. P.; Espindola, L. S., Mass spectrometry-based metabolomics approach in the isolation of bioactive natural products. *Scientific reports* **2020**, *10* (1), 1-9.
80. Lesk, A. M., *Introduction to protein architecture : the structural biology of proteins*. Oxford University Press: Oxford; New York, 2001; p xii, 347 p.
81. Lodish, H.; Berk, A.; Matsudaira, P.; Kaiser, C.; Krieger, M.; Scott, M.; Zipursky, S.; Darnell, J., *Molecular cell biology*. chapter 3: Protein structure and function. WH Freeman and Company, New York: 2004.
82. Wang, J. L.; Gao, G.; Li, Y. W.; Yang, L. Z.; Liang, Y. L.; Jin, H. Y.; Han, W. W.; Feng, Y.; Zhang, Z. M., Cloning, expression, and characterization of a thermophilic endoglucanase, AcCel12B from acidothermus cellulolyticus 11B. *International Journal of Molecular Sciences* **2015**, *16* (10), 25080-25095.
83. Zhang, Z.; Wu, S.; Stenoien, D. L.; Paša-Tolić, L., High-throughput proteomics. *Annual review of analytical chemistry* **2014**, *7*, 427-454.
84. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates III, J. R., Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews* **2013**, *113* (4), 2343-2394.
85. Shen, Y.; Zhao, R.; Belov, M. E.; Conrads, T. P.; Anderson, G. A.; Tang, K.; Paša-Tolić, L.; Veenstra, T. D.; Lipton, M. S.; Udseth, H. R., Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Analytical chemistry* **2001**, *73* (8), 1766-1775.
86. Shen, Y.; Moore, R. J.; Zhao, R.; Blonder, J.; Auberry, D. L.; Masselon, C.; Paša-Tolić, L.; Hixson, K. K.; Auberry, K. J.; Smith, R. D., High-efficiency on-line solid-phase extraction coupling to 15-150- $\mu$ m-id column liquid chromatography for proteomic analysis. *Analytical chemistry* **2003**, *75* (14), 3596-3605.

87. Shen, Y.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D., Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics. *Analytical chemistry* **2005**, *77* (10), 3090-3100.
88. Holm, M.; Joenvaara, S.; Saraswat, M.; Mustonen, H.; Tohmola, T.; Ristimäki, A.; Renkonen, R.; Haglund, C., Identification of several plasma proteins whose levels in colorectal cancer patients differ depending on outcome. *FASEB Bioadv* **2019**, *1* (12), 723-730.
89. Zhao, Y. Y.; Lin, R. C., UPLC-MS(E) application in disease biomarker discovery: the discoveries in proteomics to metabolomics. *Chem Biol Interact* **2014**, *215*, 7-16.
90. Storey, A. J.; Hardman, R. E.; Byrum, S. D.; Mackintosh, S. G.; Edmondson, R. D.; Wahls, W. P.; Tackett, A. J.; Lewis, J. A., Accurate and sensitive quantitation of the dynamic heat shock proteome using tandem mass tags. *J Proteome Res* **2020**, *19* (3), 1183-1195.
91. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **1999**, *17* (7), 676-82.
92. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001**, *19* (3), 242-7.
93. Shen, Z.; Want, E. J.; Chen, W.; Keating, W.; Nussbaumer, W.; Moore, R.; Gentle, T. M.; Siuzdak, G., Sepsis plasma protein profiling with immunodepletion, three-dimensional liquid chromatography tandem mass spectrometry, and spectrum counting. *J Proteome Res* **2006**, *5* (11), 3154-3160.
94. Nagano, K.; Taoka, M.; Yamauchi, Y.; Itagaki, C.; Shinkawa, T.; Nunomura, K.; Okamura, N.; Takahashi, N.; Izumi, T.; Isobe, T., Large-scale identification of proteins expressed in mouse embryonic stem cells. *Proteomics* **2005**, *5* (5), 1346-61.
95. Wilson, S. R.; Jankowski, M.; Pepaj, M.; Mihailova, A.; Boix, F.; Truyols, G. V.; Lundanes, E.; Greibrokk, T., 2D LC separation and determination of bradykinin in rat muscle tissue dialysate with on-line SPE-HILIC-SPE-RP-MS. *Chromatographia* **2007**, *66* (7-8), 469-474.
96. Simpson, D. C.; Ahn, S.; Pasa-Tolic, L.; Bogdanov, B.; Mottaz, H. M.; Vilkov, A. N.; Anderson, G. A.; Lipton, M. S.; Smith, R. D., Using size exclusion chromatography-RPLC and RPLC-CIEF as two-dimensional separation strategies for protein profiling. *Electrophoresis* **2006**, *27* (13), 2722-33.
97. Chen, J.; Balgley, B. M.; DeVoe, D. L.; Lee, C. S., Capillary isoelectric focusing-based multidimensional concentration/separation platform for proteome analysis. *Anal Chem* **2003**, *75* (13), 3145-52.
98. Yang, F.; Shen, Y.; Camp, D. G., 2nd; Smith, R. D., High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev Proteomics* **2012**, *9* (2), 129-34.
99. Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature chemistry* **2014**, *6* (4), 281.
100. Dear, G. J.; Munoz-Muriedas, J.; Beaumont, C.; Roberts, A.; Kirk, J.; Williams, J. P.; Campuzano, I., Sites of metabolic substitution: investigating metabolite structures utilising ion mobility and molecular modelling. *Rapid Communications in Mass Spectrometry* **2010**, *24* (21), 3157-3162.
101. McLean, J. R.; McLean, J. A.; Wu, Z.; Becker, C.; Pérez, L. M.; Pace, C. N.; Scholtz, J. M.; Russell, D. H., Factors that influence helical preferences for singly charged gas-phase peptide ions: the effects of multiple potential charge-carrying sites. *The journal of physical chemistry B* **2010**, *114* (2), 809-816.
102. Albrieux, F.; Hamidane, H. B.; Calvo, F.; Chirof, F.; Tsybin, Y. O.; Antoine, R.; Lemoine, J.; Dugourd, P., Structural preferences of gas-phase M2TMP monomers upon sequence variations. *The Journal of Physical Chemistry A* **2011**, *115* (18), 4711-4718.
103. Kaddis, C. S.; Loo, J. A., Native protein MS and ion mobility: large flying proteins with ESI. ACS Publications: 2007.
104. Loo, J. A.; Berhane, B.; Kaddis, C. S.; Wooding, K. M.; Xie, Y.; Kaufman, S. L.; Chernushevich, I. V., Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex. *Journal of the American Society for Mass Spectrometry* **2005**, *16* (7), 998-1008.
105. Duijn, E. v.; Barendregt, A.; Synowsky, S.; Versluis, C.; Heck, A. J., Chaperonin complexes monitored by ion mobility mass spectrometry. *Journal of the American Chemical Society* **2009**, *131* (4), 1452-1459.
106. Kim, S.; Pevzner, P. A., MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **2014**, *5*, 5277.
107. Diament, B. J.; Noble, W. S., Faster SEQUEST searching for peptide identification from tandem mass spectra. *J Proteome Res* **2011**, *10* (9), 3871-9.
108. Bjornson, R. D.; Carriero, N. J.; Colangelo, C.; Shifman, M.; Cheung, K. H.; Miller, P. L.; Williams, K., X!Tandem, an improved method for running X!tandem in parallel on collections of commodity computers. *J Proteome Res* **2008**, *7* (1), 293-9.

109. Brosch, M.; Yu, L.; Hubbard, T.; Choudhary, J., Accurate and sensitive peptide identification with Mascot Percolator. *J Proteome Res* **2009**, *8* (6), 3176-81.
110. Schubert, O. T.; Rost, H. L.; Collins, B. C.; Rosenberger, G.; Aebersold, R., Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc* **2017**, *12* (7), 1289-1294.
111. Mueller, L. N.; Brusniak, M.-Y.; Mani, D.; Aebersold, R., An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J Proteome Res* **2008**, *7* (01), 51-61.
112. Rajcevic, U.; Niclou, S. P.; Jimenez, C. R., Proteomics strategies for target identification and biomarker discovery in cancer. *Front Biosci* **2009**, *14* (1), 3292-303.
113. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26* (7), 966-8.
114. Zhu, W.; Smith, J. W.; Huang, C.-M., Mass spectrometry-based label-free quantitative proteomics. *BioMed Research International* **2009**, *2010*.
115. Toby, T. K.; Fornelli, L.; Kelleher, N. L., Progress in top-down proteomics and the analysis of proteoforms. *Annual review of analytical chemistry* **2016**, *9*, 499-519.
116. Catherman, A. D.; Skinner, O. S.; Kelleher, N. L., Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun* **2014**, *445* (4), 683-93.
117. Kelleher, N. L., Peer reviewed: top-down proteomics. ACS Publications: 2004.
118. R, R. J., The mechanism behind top-down UVPD experiments: making sense of apparent contradictions. *J Am Soc Mass Spectrom* **2017**, *28* (9), 1823-1826.
119. Cupp-Sutton, K. A.; Wu, S., High-throughput quantitative top-down proteomics. *Mol Omics* **2020**, *16* (2), 91-99.
120. Holt, M. V.; Wang, T.; Young, N. L., High-throughput quantitative top-down proteomics: histone H4. *Journal of The American Society for Mass Spectrometry* **2019**, *30* (12), 2548-2560.
121. Aebersold, R.; Mann, M., Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537* (7620), 347.
122. Bradshaw, R. A.; Burlingame, A. L., From proteins to proteomics. *Iubmb Life* **2005**, *57* (4-5), 267-272.
123. Ostrowski, K.; Barnard, E. A., Application of isotopically labelled specific inhibitors as a method in enzyme cytochemistry. *Exp Cell Res* **1961**, *25*, 465-8.
124. Kidd, D.; Liu, Y.; Cravatt, B. F., Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **2001**, *40* (13), 4005-15.
125. Kato, D.; Boatright, K. M.; Berger, A. B.; Nazif, T.; Blum, G.; Ryan, C.; Chehade, K. A.; Salvesen, G. S.; Bogoy, M., Activity-based probes that target diverse cysteine protease families. *Nat Chem Biol* **2005**, *1* (1), 33-8.
126. Li, Y. M.; Xu, M.; Lai, M. T.; Huang, Q.; Castro, J. L.; DiMuzio-Mower, J.; Harrison, T.; Lellis, C.; Nadin, A.; Neduveilil, J. G.; Register, R. B.; Sardana, M. K.; Shearman, M. S.; Smith, A. L.; Shi, X. P.; Yin, K. C.; Shafer, J. A.; Gardell, S. J., Photoactivated gamma-secretase inhibitors directed to the active site covalently label presenilin 1. *Nature* **2000**, *405* (6787), 689-94.
127. Sieber, S. A.; Niessen, S.; Hoover, H. S.; Cravatt, B. F., Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nat Chem Biol* **2006**, *2* (5), 274-81.
128. Patricelli, M. P.; Szardenings, A. K.; Liyanage, M.; Nomanbhoy, T. K.; Wu, M.; Weissig, H.; Aban, A.; Chun, D.; Tanner, S.; Kozarich, J. W., Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **2007**, *46* (2), 350-8.
129. Hekmat, O.; Kim, Y. W.; Williams, S. J.; He, S.; Withers, S. G., Active-site peptide "fingerprinting" of glycosidases in complex mixtures by mass spectrometry. Discovery of a novel retaining beta-1,4-glycanase in *Cellulomonas fimi*. *J Biol Chem* **2005**, *280* (42), 35126-35.
130. Salisbury, C. M.; Cravatt, B. F., Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc Natl Acad Sci USA* **2007**, *104* (4), 1171-6.
131. Adam, G. C.; Sorensen, E. J.; Cravatt, B. F., Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. *Nat Biotechnol* **2002**, *20* (8), 805-9.
132. Diether, M.; Sauer, U., Towards detecting regulatory protein-metabolite interactions. *Curr Opin Microbiol* **2017**, *39*, 16-23.
133. Lomenick, B.; Hao, R.; Jonai, N.; Chin, R. M.; Aghajan, M.; Warburton, S.; Wang, J. N.; Wu, R. P.; Gomez, F.; Loo, J. A.; Wohlschlegel, J. A.; Vondriska, T. M.; Pelletier, J.; Herschman, H. R.; Clardy, J.; Clarke, C. F.; Huang, J., Target identification using drug affinity responsive target stability (DARTS). *P Natl Acad Sci USA* **2009**, *106* (51), 21984-21989.
134. Tran, D. T.; Adhikari, J.; Fitzgerald, M. C., Stable Isotope Labeling with Amino Acids in Cell Culture

(SILAC)-based strategy for proteome-wide thermodynamic analysis of protein-ligand binding interactions. *Mol Cell Proteomics* **2014**, *13* (7), 1800-1813.

135. Piazza, I.; Kochanowski, K.; Cappelletti, V.; Fuhrer, T.; Noor, E.; Sauer, U.; Picotti, P., A map of protein-metabolite interactions reveals principles of chemical communication. *Cell* **2018**, *172* (1-2), 358-372 e23.

136. Franken, H.; Mathieson, T.; Childs, D.; Sweetman, G. M.; Werner, T.; Togel, I.; Doce, C.; Gade, S.; Bantscheff, M.; Drewes, G.; Reinhard, F. B.; Huber, W.; Savitski, M. M., Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat Protoc* **2015**, *10* (10), 1567-93.

137. Mateus, A.; Maatta, T. A.; Savitski, M. M., Thermal proteome profiling: unbiased assessment of protein state through heat-induced stability changes. *Proteome Sci* **2016**, *15*, 13.

138. Savitski, M. M.; Reinhard, F. B.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Molina, D. M.; Jafari, R.; Dovega, R. B.; Kläeger, S., Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **2014**, *346* (6205), 1255784.

139. Childs, D.; Bach, K.; Franken, H.; Anders, S.; Kurzawa, N.; Bantscheff, M.; Savitski, M. M.; Huber, W., Nonparametric analysis of thermal proteome profiles reveals novel drug-binding proteins. *Mol Cell Proteomics* **2019**, *18* (12), 2506-2515.

140. Mateus, A.; Kurzawa, N.; Becher, I.; Sridharan, S.; Helm, D.; Stein, F.; Typas, A.; Savitski, M. M., Thermal proteome profiling for interrogating protein interactions. *Molecular Systems Biology* **2020**, *16* (3), e9232.

141. Sridharan, S.; Kurzawa, N.; Werner, T.; Günthner, I.; Helm, D.; Huber, W.; Bantscheff, M.; Savitski, M. M., Proteome-wide solubility and thermal stability profiling reveals distinct regulatory roles for ATP. *Nat Commun* **2019**, *10* (1), 1-13.

142. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod* **2016**, *79* (3), 629-661.

143. Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J., The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* **2015**, *14* (2), 111-29.

144. Lobanovska, M.; Pilla, G., Penicillin's discovery and antibiotic resistance: lessons for the future? *Yale J Biol Med* **2017**, *90* (1), 135-145.

145. Weaver, B. A., How Taxol/paclitaxel kills cancer cells. *Mol Biol Cell* **2014**, *25* (18), 2677-2681.

146. Su, X. Z.; Miller, L. H., The discovery of artemisinin and the Nobel Prize in Physiology or Medicine. *Sci China Life Sci* **2015**, *58* (11), 1175-9.

147. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J Nat Prod* **2020**.

148. Prior, M.; Chiruta, C.; Currais, A.; Goldberg, J.; Ramsey, J.; Dargusch, R.; Maher, P. A.; Schubert, D., Back to the future with phenotypic screening. *ACS Chem Neurosci* **2014**, *5* (7), 503-13.

149. Weller, M. G., A unifying review of bioassay-guided fractionation, effect-directed analysis and related techniques. *Sensors (Basel)* **2012**, *12* (7), 9181-209.

150. Bergsdorf, C.; Ottl, J., Affinity-based screening techniques: their impact and benefit to increase the number of high quality leads. *Expert Opin Drug Discov* **2010**, *5* (11), 1095-107.

151. Rinschen, M. M.; Ivanisevic, J.; Giera, M.; Siuzdak, G., Identification of bioactive metabolites using activity metabolomics. *Nat Rev Mol Cell Biol* **2019**, *20* (6), 353-367.

152. Zhuo, R.; Liu, H.; Liu, N.; Wang, Y., Ligand Fishing: A remarkable strategy for discovering bioactive compounds from complex mixture of natural products. *Molecules* **2016**, *21* (11).

153. Fu, Y.; Mo, H. Y.; Gao, W.; Hong, J. Y.; Lu, J.; Li, P.; Chen, J., Affinity selection-based two-dimensional chromatography coupled with high-performance liquid chromatography-mass spectrometry for discovering xanthine oxidase inhibitors from Radix Salviae Miltiorrhizae. *Anal Bioanal Chem* **2014**, *406* (20), 4987-95.

154. Munigunti, R.; Mulabagal, V.; Calderon, A. I., Screening of natural compounds for ligands to PfTrxR by ultrafiltration and LC-MS based binding assay. *J Pharm Biomed Anal* **2011**, *55* (2), 265-71.

155. Yang, X. X.; Xu, F.; Wang, D.; Yang, Z. W.; Tan, H. R.; Shang, M. Y.; Wang, X.; Cai, S. Q., Development of a mitochondria-based centrifugal ultrafiltration/liquid chromatography/mass spectrometry method for screening mitochondria-targeted bioactive constituents from complex matrixes: herbal medicines as a case study. *J Chromatogr A* **2015**, *1413*, 33-46.

156. Cen, Y.; Xiao, A.; Chen, X.; Liu, L., Isolation of alpha-amylase inhibitors from Kadsura longipedunculata using a high-speed counter-current chromatography target guided by centrifugal ultrafiltration with LC-MS. *Molecules* **2016**, *21* (9).

157. Li, H.; Song, F.; Xing, J.; Tsao, R.; Liu, Z.; Liu, S., Screening and structural characterization of alpha-glucosidase inhibitors from hawthorn leaf flavonoids extract by ultrafiltration LC-DAD-MS(n) and SORI-CID FTICR

- MS. *J Am Soc Mass Spectrom* **2009**, *20* (8), 1496-503.
158. Choi, Y.; Jung, Y.; Kim, S. N., Identification of eupatilin from artemisia argyi as a selective PPARalpha agonist using affinity selection ultrafiltration LC-MS. *Molecules* **2015**, *20* (8), 13753-63.
159. Choi, Y.; van Breemen, R. B., Development of a screening assay for ligands to the estrogen receptor based on magnetic microparticles and LC-MS. *Comb Chem High Throughput Screen* **2008**, *11* (1), 1-6.
160. Rush, M. D.; Walker, E. M.; Burton, T.; van Breemen, R. B., Magnetic Microbead Affinity Selection Screening (MagMASS) of botanical extracts for inhibitors of 15-lipoxygenase. *J Nat Prod* **2016**, *79* (11), 2898-2902.
161. Marszall, M. P.; Bucinski, A., A protein-coated magnetic beads as a tool for the rapid drug-protein binding study. *J Pharmaceut Biomed* **2010**, *52* (3), 420-424.
162. Wubshet, S. G.; Brighente, I. M. C.; Moaddel, R.; Staerk, D., Magnetic ligand fishing as a targeting tool for HPLC-HRMS-SPE-NMR: alpha-glucosidase inhibitory ligands and alkylresorcinol glycosides from eugenia catharinae. *J Nat Prod* **2015**, *78* (11), 2657-2665.
163. Li, F.; Zhang, Y. M.; Qiu, D. Y.; Kang, J. W., Screening of epidermal growth factor receptor inhibitors in natural products by capillary electrophoresis combined with high performance liquid chromatography-tandem mass spectrometry. *J Chromatogr A* **2015**, *1400*, 117-123.
164. Hou, G. Y.; Niu, J.; Song, F. R.; Liu, Z. Q.; Liu, S. Y., Studies on the interactions between ginsenosides and liposome by equilibrium dialysis combined with ultrahigh performance liquid chromatography-tandem mass spectrometry. *J Chromatogr B* **2013**, *923*, 1-7.
165. Song, H. P.; Chen, J.; Hong, J. Y.; Hao, H. P.; Qi, L. W.; Lu, J.; Fu, Y.; Wu, B.; Yang, H.; Li, P., A strategy for screening of high-quality enzyme inhibitors from herbal medicines based on ultrafiltration LC-MS and in silico molecular docking. *Chem Commun* **2015**, *51* (8), 1494-1497.
166. Rycenga, H. B.; Long, D. T., The evolving role of DNA inter-strand crosslinks in chemotherapy. *Curr Opin Pharmacol* **2018**, *41*, 20-26.
167. Dasari, S.; Tchounwou, P. B., Cisplatin in cancer therapy: molecular mechanisms of action. *Eur J Pharmacol* **2014**, *740*, 364-378.
168. Chanes, R. E.; Condit, P. T.; Bottomley, R. H.; Nisimblat, W., Combined actinomycin D and vincristine in the treatment of patients with cancer. *Cancer* **1971**, *27* (3), 613-7.
169. Gurova, K., New hopes from old drugs: revisiting DNA-binding small molecules as anticancer agents. *Future Oncol* **2009**, *5* (10), 1685-1704.
170. Hurley, L. H., DNA and its associated processes as targets for cancer therapy. *Nat Rev Cancer* **2002**, *2* (3), 188-200.
171. Wang, M.; Yu, Y.; Liang, C.; Lu, A.; Zhang, G., Recent advances in developing small molecules targeting nucleic acid. *Int J Mol Sci* **2016**, *17* (6).
172. Xu, N.; Yang, H.; Cui, M.; Wan, C.; Liu, S., High-performance liquid chromatography-electrospray ionization-mass spectrometry ligand fishing assay: a method for screening triplex DNA binders from natural plant extracts. *Anal Chem* **2012**, *84* (5), 2562-8.
173. Hage, D. S.; Anguizola, J. A.; Bi, C.; Li, R.; Matsuda, R.; Papastavros, E.; Pfaunmiller, E.; Vargas, J.; Zheng, X., Pharmaceutical and biomedical applications of affinity chromatography: recent trends and developments. *J Pharm Biomed Anal* **2012**, *69*, 93-105.
174. Nguyen, T. D.; Lesani, M.; Forrest, I.; Lan, Y.; Dean, D. A.; Gibaut, Q. M. R.; Guo, Y.; Hossain, E.; Olvera, M.; Panlilio, H.; Parab, A. R.; Wu, C.; Bernatchez, J. A.; Cichewicz, R. H.; McCall, L. I., Local phenomena shape backyard soil metabolite composition. *Metabolites* **2020**, *10* (3).
175. Liu, X. F.; Xiang, L. M.; Zhou, Q.; Carralot, J. P.; Prunotto, M.; Niederfellner, G.; Pastan, I., Actinomycin D enhances killing of cancer cells by immunotoxin RG7787 through activation of the extrinsic pathway of apoptosis. *P Natl Acad Sci USA* **2016**, *113* (38), 10666-10671.
176. Sobell, H. M.; Jain, S. C., Stereochemistry of Actinomycin Binding to DNA .2. Detailed molecular model of actinomycin-DNA complex and its implications. *J Mol Biol* **1972**, *68* (1), 21-+.
177. Kamitori, S.; Takusagawa, F., Crystal-structure of the 2/1 complex between D (Gaagcttc) and the anticancer drug actinomycin-D. *J Mol Biol* **1992**, *225* (2), 445-456.
178. Zhang, X. F.; Ye, X. W.; Chai, W. Y.; Lian, X. Y.; Zhang, Z. Z., New metabolites and bioactive actinomycins from marine-derived streptomyces sp ZZ338. *Mar Drugs* **2016**, *14* (10).
179. Li, X. L.; Hu, Y. J.; Wang, H.; Yu, B. Q.; Yue, H. L., Molecular spectroscopy evidence of berberine binding to DNA: comparative binding and thermodynamic profile of intercalation. *Biomacromolecules* **2012**, *13* (3), 873-880.
180. Kamath, S.; Skeels, M.; Pai, A., Significant differences in alkaloid content of *Coptis chinensis* (Huanglian), from its related american species. *Chin Med* **2009**, *4*, 17.
181. Jung, H. A.; Yoon, N. Y.; Bae, H. J.; Min, B. S.; Choi, J. S., Inhibitory activities of the alkaloids from *Coptidis*

- rhizoma against aldose reductase. *Arch Pharm Res* **2008**, *31* (11), 1405-1412.
182. Dumont, E.; Monari, A., Interaction of palmatine with DNA: an environmentally controlled phototherapy drug. *J Phys Chem B* **2015**, *119* (2), 410-419.
183. Papi, F.; Ferraroni, M.; Rigo, R.; Da Ros, S.; Bazzicalupi, C.; Sissi, C.; Gratteri, P., Role of the benzodioxole group in the interactions between the natural alkaloids chelerythrine and coptisine and the human telomeric G-quadruplex DNA. A multiapproach investigation. *J Nat Prod* **2017**, *80* (12), 3129-3136.
184. He, M.; Grkovic, T.; Evans, J. R.; Thornburg, C. C.; Akee, R. K.; Thompson, J. R.; Whitt, J. A.; Harris, M. J.; Loyal, J. A.; Britt, J. R.; Jia, L.; White, J. D.; Newman, D. J.; O'Keefe, B. R., The NCI library of traditional Chinese medicinal plant extracts-Preliminary assessment of the NCI-60 activity and chemical profiling of selected species. *Fitoterapia* **2019**, *137*, 104285.
185. Yang, H.; Wang, Y.; Yu, W.; Shi, L.; Wang, H.; Su, R.; Chen, C.; Liu, S., Screening and investigation of triplex DNA binders from *Stephania tetrandra* S. Moore by a combination of peak area-fading ultra high-performance liquid chromatography with Orbitrap mass spectrometry and optical spectroscopies. *Journal of separation science* **2018**, *41* (14), 2878-2885.
186. Xie, Z.; Xu, X.; Xie, C.; Liang, Z.; Yang, M.; Huang, J.; Yang, D., Preparative isolation of tetrandrine and fangchinoline from *Radix Stephania tetrandra* using reversed-phase flash chromatography. *Journal of Liquid Chromatography & Related Technologies* **2014**, *37* (3), 343-352.
187. Li, H.; Chen, X.; Zhou, S.-J., Dauricine combined with clindamycin inhibits severe pneumonia co-infected by influenza virus H5N1 and *Streptococcus pneumoniae* in vitro and in vivo through NF- $\kappa$ B signaling pathway. *Journal of pharmacological sciences* **2018**, *137* (1), 12-19.
188. Wu, M.-Y.; Wang, S.-F.; Cai, C.-Z.; Tan, J.-Q.; Li, M.; Lu, J.-J.; Chen, X.-P.; Wang, Y.-T.; Zheng, W.; Lu, J.-H., Natural autophagy blockers, dauricine (DAC) and daurisolone (DAS), sensitize cancer cells to camptothecin-induced toxicity. *Oncotarget* **2017**, *8* (44), 77673.
189. Luo, H.; Peng, M.; Ye, H.; Chen, L.; Peng, A.; Tang, M.; Zhang, F.; Shi, J., Predictable and linear scale-up of four phenolic alkaloids separation from the roots of *Menispermum dauricum* using high-performance counter-current chromatography. *J Chromatogr B Analyt Technol Biomed Life Sci* **2010**, *878* (22), 1929-33.
190. Fournet, A.; Cavé, A.; Duté, P.; Weber, J.-F.; Bruneton, J., Bis-benzylisoquinoline alkaloids from *abuta pahnii*. *Phytochemistry* **1987**, *26* (7), 2136-2137.
191. Schiff, P. L., Jr., Bisbenzylisoquinoline alkaloids. *J Nat Prod* **1991**, *54* (3), 645-749.
192. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycomics. *Nucleic Acids Research* **2009**, *37*, D233-D238.
193. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **2014**, *42* (Database issue), D490-5.
194. Linger, J. G.; Adney, W. S.; Darzins, A., Heterologous expression and extracellular secretion of cellulolytic enzymes by *Zymomonas mobilis*. *Appl Environ Microb* **2010**, *76* (19), 6360-6369.
195. Wu, W. H.; Hung, W. C.; Lo, K. Y.; Chen, Y. H.; Wan, H. P.; Cheng, K. C., Bioethanol production from taro waste using thermo-tolerant yeast *Kluyveromyces marxianus* K21. *Bioresource Technol* **2016**, *201*, 27-32.
196. Jorgensen, H.; Eriksson, T.; Borjesson, J.; Tjerneld, F.; Olsson, L., Purification and characterization of five cellulases and one xylanase from *Penicillium brasilianum* IBT 20888. *Enzyme Microb Tech* **2003**, *32* (7), 851-861.
197. Adav, S. S.; Li, A. A.; Manavalan, A.; Punt, P.; Sze, S. K., Quantitative iTRAQ secretome analysis of *Aspergillus niger* reveals novel hydrolytic enzymes. *J Proteome Res* **2010**, *9* (8), 3932-3940.
198. Chauvigne-Hines, L. M.; Anderson, L. N.; Weaver, H. M.; Brown, J. N.; Koech, P. K.; Nicora, C. D.; Hofstad, B. A.; Smith, R. D.; Wilkins, M. J.; Callister, S. J.; Wright, A. T., Suite of activity-based probes for cellulose-degrading enzymes. *J Am Chem Soc* **2012**, *134* (50), 20521-32.
199. Li, N.; Overkleeft, H. S.; Florea, B. I., Activity-based protein profiling: an enabling technology in chemical biology research. *Curr Opin Chem Biol* **2012**, *16* (1-2), 227-33.
200. Anderson, L. N.; Culley, D. E.; Hofstad, B. A.; Chauvigne-Hines, L. M.; Zink, E. M.; Purvine, S. O.; Smith, R. D.; Callister, S. J.; Magnuson, J. M.; Wright, A. T., Activity-based protein profiling of secreted cellulolytic enzyme activity dynamics in *Trichoderma reesei* QM6a, NG14, and RUT-C30. *Mol Biosyst* **2013**, *9* (12), 2992-3000.
201. Chundawat, S. P.; Lipton, M. S.; Purvine, S. O.; Uppugundla, N.; Gao, D.; Balan, V.; Dale, B. E., Proteomics-based compositional analysis of complex cellulase-hemicellulase mixtures. *J Proteome Res* **2011**, *10* (10), 4365-72.
202. Jansen, R. S.; Addie, R.; Merckx, R.; Fish, A.; Mahakena, S.; Bleijerveld, O. B.; Altelaar, M.; IJlst, L.; Wanders, R. J.; Borst, P.; van de Wetering, K., N-lactoyl-amino acids are ubiquitous metabolites that originate from CNDP2-mediated reverse proteolysis of lactate and amino acids. *P Natl Acad Sci USA* **2015**, *112* (21), 6601-6606.
203. Richardson, S.; Gorton, L., Characterisation of the substituent distribution in starch and cellulose derivatives.

*Analytica Chimica Acta* **2003**, 497 (1-2), 27-65.

204. Chundawat, S. P. S.; Lipton, M. S.; Purvine, S. O.; Uppugundla, N.; Gao, D. H.; Balan, V.; Dale, B. E., Proteomics-based compositional analysis of complex cellulase-hemicellulase mixtures. *J Proteome Res* **2011**, 10 (10), 4365-4372.
205. Man, J. M.; Yang, Y.; Huang, J.; Zhang, C. Q.; Zhang, F. M.; Wang, Y. P.; Gu, M. H.; Liu, Q. Q.; Wei, C. X., Morphology and structural properties of high-amylose rice starch residues hydrolysed by amyloglucosidase. *Food Chem* **2013**, 138 (4), 2089-2098.
206. Marchler-Bauer, A.; Derbyshire, M. K.; Gonzales, N. R.; Lu, S. N.; Chitsaz, F.; Geer, L. Y.; Geer, R. C.; He, J.; Gwadz, M.; Hurwitz, D. I.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Wang, Z. X.; Yamashita, R. A.; Zhang, D. C.; Zheng, C. J.; Bryant, S. H., CDD: NCBI's conserved domain database. *Nucleic Acids Research* **2015**, 43 (D1), D222-D226.
207. Jordan, D. B.; Bowman, M. J.; Braker, J. D.; Dien, B. S.; Hector, R. E.; Lee, C. C.; Mertens, J. A.; Wagschal, K., Plant cell walls to ethanol. *Biochem J* **2012**, 442 (2), 241-52.
208. Qu, Y.; Feng, J.; Deng, S.; Cao, L.; Zhang, Q. B.; Zhao, R.; Zhang, Z. R.; Jiang, Y. X.; Zink, E. M.; Baker, S. E.; Lipton, M. S.; Pasa-Tolic, L.; Hu, J. Z.; Wu, S., Structural analysis of N- and O-glycans using ZIC-HILIC/dialysis coupled to NMR detection. *Fungal Genet Biol* **2014**, 72, 207-215.
209. Ghose, T. K., Measurement of cellulase activities. *Pure and Applied Chemistry* **1987**, 59 (2), 257-268.
210. Ooshima, H.; Burns, D. S.; Converse, A. O., Adsorption of cellulase from trichoderma-reesei on cellulose and lignocellulosic residue in wood pretreated by dilute sulfuric-acid with explosive decompression. *Biotechnol Bioeng* **1990**, 36 (5), 446-452.
211. Zhang, Z. R.; Wu, S.; Stenoien, D. L.; Pasa-Tolic, L., High-throughput proteomics. *Annu Rev Anal Chem* **2014**, 7, 427-454.
212. Kim, S.; Pevzner, P. A., MS-GF plus makes progress towards a universal database search tool for proteomics. *Nature Communications* **2014**, 5.

## Appendix

### Supporting Information for Chapter 3

#### List of Contents

<b>Table S1.</b> List of 62 herbs.....	83
<b>Table S2.</b> <sup>1</sup> H NMR (600 MHz) data for actinomycin D ( <b>9</b> ) and V ( <b>10</b> ) in CDCl <sub>3</sub> .....	84
<b>Table S3.</b> <sup>13</sup> C NMR (150 MHz) data for actinomycin D ( <b>9</b> ) and V ( <b>10</b> ) in CDCl <sub>3</sub> .....	85
<b>Table S4.</b> <sup>1</sup> H NMR (600 MHz) data for actinomycin X <sub>0β</sub> ( <b>11</b> ) in CDCl <sub>3</sub> .....	86
<b>Table S5.</b> <sup>1</sup> H NMR (600 MHz), and <sup>13</sup> C NMR (150 MHz) data for berberine ( <b>12</b> ) in MeOH- <i>d</i> <sub>4</sub> .....	87
<b>Table S6.</b> <sup>1</sup> H NMR (600 MHz), and <sup>13</sup> C NMR (150 MHz) data for palmatine ( <b>13</b> ) in MeOH- <i>d</i> <sub>4</sub> .....	88
<b>Table S7.</b> <sup>1</sup> H NMR (600 MHz), and <sup>13</sup> C NMR (150 MHz) data for coptisine ( <b>14</b> ) in MeOH- <i>d</i> <sub>4</sub> .....	89
<b>Figure S1.</b> Overflow of DNA-ligand complex dissociation .....	90
<b>Figure S2.</b> Calibration curve analysis (UV peak area) of 4 intercalators and 1 groove binder.....	91
<b>Figure S3.</b> EIC analysis of the 4 individual intercalators in the assay.....	92
<b>Figure S4.</b> EIC analysis of bisbenzimidazole (H33258), melphalan, and 4 intercalators mixture.....	93
<b>Figure S5.</b> Annotation and mirror plots of actinomycin D and its analogues using GNPS.....	94
<b>Figure S6.</b> Results of the high throughput LLAMAS development.....	95
<b>Figure S7.</b> Confirmation of the DNA binding activities of fangchinoline ( <b>15</b> ), tetrandrine ( <b>16</b> ), daurisolone ( <b>17</b> ), and dauricine ( <b>18</b> ).....	96
<b>Figure S8.</b> <sup>1</sup> H NMR (600 MHz) spectrum of actinomycin D ( <b>9</b> ) in CDCl <sub>3</sub> .....	97
<b>Figure S9.</b> <sup>13</sup> C NMR (150 MHz) spectrum of actinomycin D ( <b>9</b> ) in CDCl <sub>3</sub> .....	98
<b>Figure S10.</b> <sup>1</sup> H NMR (600 MHz) spectrum of actinomycin V ( <b>10</b> ) in CDCl <sub>3</sub> .....	99
<b>Figure S11.</b> <sup>13</sup> C NMR (150 MHz) spectrum of actinomycin V ( <b>10</b> ) in CDCl <sub>3</sub> .....	100
<b>Figure S12.</b> <sup>1</sup> H NMR (600 MHz) spectrum of actinomycin X <sub>0β</sub> ( <b>11</b> ) in CDCl <sub>3</sub> .....	101
<b>Figure S13.</b> <sup>1</sup> H NMR (600 MHz) spectrum of berberine ( <b>12</b> ) in MeOH- <i>d</i> <sub>4</sub> .....	102
<b>Figure S14.</b> <sup>13</sup> C NMR (150 MHz) spectrum of berberine ( <b>12</b> ) in MeOH- <i>d</i> <sub>4</sub> .....	103
<b>Figure S15.</b> <sup>1</sup> H NMR (600 MHz) spectrum of palmatine ( <b>13</b> ) in MeOH- <i>d</i> <sub>4</sub> .....	104
<b>Figure S16.</b> <sup>13</sup> C NMR (150 MHz) spectrum of palmatine ( <b>13</b> ) in MeOH- <i>d</i> <sub>4</sub> .....	105
<b>Figure S17.</b> <sup>1</sup> H NMR (600 MHz) spectrum of coptisine ( <b>14</b> ) in MeOH- <i>d</i> <sub>4</sub> .....	106
<b>Figure S18.</b> <sup>13</sup> C NMR (150 MHz) spectrum of coptisine ( <b>14</b> ) in MeOH- <i>d</i> <sub>4</sub> .....	107
<b>Figure S19.</b> <sup>1</sup> H NMR (500 MHz) spectrum of fangchinoline ( <b>15</b> ) in CDCl <sub>3</sub> .....	108
<b>Figure S20.</b> <sup>1</sup> H NMR (500 MHz) spectrum of tetrandrine ( <b>16</b> ) in CDCl <sub>3</sub> .....	109
<b>Figure S21.</b> <sup>1</sup> H NMR (500 MHz) spectrum of daurisolone ( <b>17</b> ) in CDCl <sub>3</sub> .....	110
<b>Figure S22.</b> <sup>1</sup> H NMR (500 MHz) spectrum of dauricine ( <b>18</b> ) in CDCl <sub>3</sub> .....	111



**Table S1.** List of 62 herbs.

Sample #	Name	Extension
1	Vetivert root	<i>Vetiveria zizanoides</i> , conventionally grown roots
2	Canada Snakeroot	<i>Asarum canadense</i> , wildcrafted
3	Atractylodes	<i>Atractylodes macrocephala</i> , sulfur free root
4	Schizandra berries	<i>Schizandrae chinensis</i> , exrta fine grade, organic
5	Stone Root	<i>Collinsonia canadensis</i> , conventionally grown root
6	Elecampane	<i>Inula helinum</i> , organic root
7	Ladies Mantle	<i>Alchemilla vulgaris</i> , organic, aerial parts
8	Butterbur	<i>Petasites frigidus</i> , root
9	Dan Shen	<i>Salvia miltiorrhiza</i> , sulphur free
10	Osha root	<i>Ligusticum porterii</i> , organic whole roots
11	Cinnamon Bark Sticks	<i>Cinnamomum cassia</i> , organic
12	Jujube Dates	<i>Ziziphus jujube</i> , sulfur free fruits
13	Rehmannia root raw	<i>Rehmanniae glutinosa</i> , sulphur free
14	Muir Puama	<i>Ptychopetalum dacordes</i> , wildcrafted
15	Kava root	<i>Piper methysticum</i> , grown without pesticides
16	Fo Ti (He Shou Wu)	<i>Polygonum multiflorum</i> , sulfur free root
17	Valerian root	<i>Valeriana officinalis</i> , organic
18	Blue Malva	<i>Malva officinalis</i> , dried flowers
19	Boldo	<i>Pemus boldus</i> , leaves
20	Alkanet Root	<i>Alkanna tinctorial</i> , chopped root
21	Honeysuckle Flowers	<i>Lonicera japonica</i> , sulfur free
22	California Poppy	<i>Escholtzia californica</i> , organic, aerial portions
23	Licorice Chinese	<i>Glycerrhiza uralensis</i> , sulfur free
24	Chrysanthemum flowers	<i>Chrysanthemum morifolium</i> , sulfur free flowers
25	Horney Goat Weed	<i>Epimedium brevicornum max.</i> , sulfur free herb
26	Chaparral	<i>Larrea divaricata</i> , wild crafted leaves
27	Oakmoss	<i>Evernia furfuracea</i> , lychen, not for internal use
28	Dang Quai	<i>Angelica chinensis</i> , sulfur free root slices
29	Wakame	<i>Alaria marginita</i> , wildcrafted seavegetable
30	Astragalus	<i>Astragalus membranaceous</i> , organic, root
31	Mangolia bark	<i>Magnolia officinalis</i> , sulfur free
32	Bai Zhi	<i>Angelica dahurica</i> , sulfur free, root
33	Peony Root	<i>Paeonia lactiflora</i> , sulfur free
34	Chuan Xin Lian	<i>Andrographis paniculata</i> , sulfur free herb
35	Poke Root	<i>Phytolacca americana</i> , wildcrafted
36	Ajowan (Ajwan) Seed	<i>Apium graveolens</i> , seeds, whole
37	Deer's Tongue Leaf	<i>Trilisa odoratissima</i>
38	Kelp	<i>Neriocystis sp.</i> , wildcrafted
39	Eleuthero	<i>Eleutherococcus senticosus</i> , organic root
40	Gold Thread (Coptis)	<i>Coptis chinensis</i> , root slices
41	Bupleurum	<i>Bupleurum chinensis</i> , sulfur free roots
42	Chaste Tree Berry	<i>Vitex agnus-castus</i> , organic
43	Codonopsis	<i>Codonopsis pilosula</i> , sulfur free root
44	Lychii berries	<i>Lycium chinensis</i> , sulfur free
45	Calamus root	<i>Acorus calamus</i> , cut sifted
46	Quassia	<i>Picrassa famara</i>
47	Orris Root	<i>Iris germanica var. Florentina</i>
48	Papaya Leaf	<i>Carica papaya</i> , conventionally grown leaf
49	Angelica	<i>Angelica archangelica</i> , organic root
50	Tansy flowering tops	<i>Tanacetum vulgare</i> , wildcrafted
51	Carydalis root	<i>Corydalis yanhouso</i> , sulfur free
52	Isatis	<i>Isatidis/Baphicanthus</i> , root
53	Pau d'Arco	<i>Tabebuia impetiginosa</i> , conventionally grown, bark
54	Pygeum	<i>Pygeum africanus</i> , bark powder
55	Gymnema sylvestris	leaf, cut/sift
56	Tonka beans whole	<i>Dipteryx odorata</i>
57	Chuan xiong	<i>Ligusticum walichii</i> , sulfur free root
58	Blue Flag	<i>Iris versicolor</i> , rhizome, cut/sift
59	Job's tears	<i>Coix lachryma-jobi</i> , sulfur free
60	Skullcap, Chinese	<i>Scutellaria baicalensis</i> , sulfur free root
61	Reishi mushroom	<i>Ganoderma lucidum</i> , sulphur free
62	High John the Conqueror	<i>Ipomea jalapa</i> , root

**Table S2.**  $^1\text{H}$  NMR (600 MHz) data for actinomycin D (**9**) and V (**10**) in  $\text{CDCl}_3$ 

No.	<b>9</b>	<b>10</b>
<b>3</b>	7.59, d (7.7)	7.59, d (7.7)
<b>4</b>	7.35, d (7.7)	7.36, d (7.7)
<b>2'</b>	4.51, dd (7.1, 2.5)	4.58, m
<b>4'</b>	3.53, dd (9.5, 6.0)	3.59, m
<b>5'</b>	2.16, m	2.13, m
<b>6'</b>	1.09, d (6.6) <sup>b</sup>	1.13, d (6.8) <sup>b</sup>
<b>7'</b>	0.85, d (6.6) <sup>b</sup>	0.91, d (6.8) <sup>b</sup>
<b>9'</b>	3.70, m; 3.95, m	3.98, d (19.5); 4.54, m
<b>10'</b>	2.07, 2.23, m	--
<b>11'</b>	1.80, m; 2.61, m	2.34, d (17.5); 3.65 d (17.5)
<b>12'</b>	5.98, d (9.2)	6.56, d (10.0)
<b>14'</b>	2.85, s	2.90, s <sup>d</sup>
<b>15'</b>	3.63, d (17.5); 4.71, d (17.5)	3.71, m; 4.59, m
<b>17'</b>	2.91, s	2.93, s
<b>18'</b>	2.67, d (9.4)	2.66, m <sup>c</sup>
<b>19'</b>	2.61, m	2.64, m
<b>20'</b>	0.72, d (6.5) <sup>a</sup>	0.75, d (6.0) <sup>a</sup>
<b>21'</b>	0.92, d (6.5) <sup>a</sup>	0.95, d (6.0) <sup>a</sup>
<b>23'</b>	5.19, dd (6.5, 2.4)	5.25, dd (5.6, 2.2)
<b>24'</b>	1.22, d (6.3)	1.26, d (6.7)
<b>NH-2'</b>	7.15, d (7.0)	7.17, d (7.3)
<b>NH-4'</b>	8.19, d (6.0)	7.76, d (5.8)
<b>2''</b>	4.61, dd (6.7, 2.5)	4.52, m
<b>4''</b>	3.56, dd (10.0, 6.0)	3.70, m
<b>5''</b>	2.12, m	2.23, m
<b>6''</b>	1.08, d (6.6) <sup>b</sup>	1.15, d (6.8) <sup>b</sup>
<b>7''</b>	0.87, d (6.6) <sup>b</sup>	0.90, d (6.8) <sup>b</sup>
<b>9''</b>	3.67, m; 3.81, m	3.73, m; 3.91m
<b>10''</b>	2.07, 2.23, m	2.21, m; 2.27, m
<b>11''</b>	1.84, m; 2.90, m	1.86, m; another H not detected
<b>12''</b>	5.9, d (9.2)	5.94, d (9.2)
<b>14''</b>	2.85, s	2.89, s <sup>d</sup>
<b>15''</b>	3.60, d (17.5); 4.79, d (17.5)	3.65, d (17.5); 4.71, d (17.5)
<b>17''</b>	2.90, s	2.91, s
<b>18''</b>	2.67, d (9.4)	2.71, m <sup>c</sup>
<b>19''</b>	2.61, m	2.63, m
<b>20''</b>	0.72, d (6.5) <sup>a</sup>	0.74, d (6.0) <sup>a</sup>
<b>21''</b>	0.94, d (6.5) <sup>a</sup>	0.98, d (6.0) <sup>a</sup>
<b>23''</b>	5.15, dd (6.5, 2.4)	5.17, dd (6.2, 2.2)
<b>24''</b>	1.22, d (6.3)	1.12, d (6.6)
<b>NH-2''</b>	7.7, d (6.6)	7.68, d (5.8)
<b>NH-4''</b>	8.01, d (6.2)	8.23, d (6.0)

<sup>a-d</sup> The data with the same labels in each column may be interchanged.

**Table S3.**  $^{13}\text{C}$  NMR (150 MHz) data for actinomycin D (**9**) and V (**10**) in  $\text{CDCl}_3$ 

No.	<b>9</b>	<b>10</b>	No.	<b>9</b>	<b>10</b>
1	129.2	129.3	17'	39.3	39.5
2	132.5	132.2	18''	71.2	71.3
3	125.8	126.2	18'	71.4	71.6
4	130.4	130.5	19''	27.0	27.2
5	127.8	128.1	19'	27.1	27.1
6	140.6	140.7	2''	54.9	55.1
7	145.2	145.2	2'	55.2	54.9
8	113.6	113.8	20''	19.1	19.2
9	179.1	179.2	20'	19.2	19.2
10	147.7	147.5	21''	21.7	21.9
11	101.7	101.9	21'	21.6	21.7
12	145.9	146.1	22''	167.8	167.6
13	15.1	15.2	22'	167.7	167.6
14	7.8	7.9	23''	75.1	74.9
1''	166.6	166.4	23'	75.0	74.8
1'	166.7	166.4	24''	17.8	17.8
10''	22.9	23.1	24'	17.4	17.3
10'	23.1	208.9	3''	169.1	169.2
11''	31.4	31.2	3'	168.7	169.1
11'	31.0	42.0	4''	58.8	57.4
12''	56.6	56.7	4'	58.9	58.8
12'	56.4	54.5	5''	31.9	32.1
13''	173.4	173.7	5'	31.6	31.9
13'	173.4	172.9	6''	19.1	19.1
14''	35.0	35.2	6'	19.1	19.0
14'	35.1	35.0	7''	19.3	19.4
15''	51.4	51.5	7'	19.4	19.3
15'	51.4	51.5	8''	173.7	173.5
16''	166.7	166.7	8'	173.3	174.2
16'	166.4	166.0	9''	47.7	47.6
17''	39.2	39.2	9'	47.4	53.0

**Table S4.**  $^1\text{H}$  NMR (600 MHz) data for actinomycin X<sub>0β</sub> (**11**) in CDCl<sub>3</sub>

No.	$\delta_{\text{H}}$ ( $J$ in Hz)
3	7.65, d (7.7)
4	7.35, d (7.7)
2'	4.83, dd (6.2, 2.2)
4'	3.57, m
6'	1.12, d (6.7) <sup>d</sup>
7'	0.88, d (6.7) <sup>d</sup>
12'	6.07, dd (9.3, 2.7)
14'	2.88, s
15'	3.60, m; 4.54, d (17.6)
17'	2.93, s <sup>a</sup>
18'	2.68, (9.1)
19'	2.66, m
20'	0.74, d (6.7) <sup>b</sup>
21'	0.96, d (6.7) <sup>b</sup>
23'	5.25, m
24'	1.29, d (6.2) <sup>c</sup>
NH-2'	7.46, d (6.2)
NH-4'	7.51, d (6.9)
2''	4.5, d (6.2, 2.5)
4''	3.74, t
6''	1.14, d (6.7) <sup>d</sup>
7''	0.91, d (6.7) <sup>d</sup>
12''	5.99, d (9.2)
14''	2.88, s
15''	3.66, d (17.2); 4.73, d (17.2)
17''	2.94, s <sup>a</sup>
18''	2.72, (9.3)
19''	2.66, m
20''	0.75, d (6.7) <sup>b</sup>
21''	0.97, d (6.7) <sup>b</sup>
23''	5.25, m
24''	1.26, d (6.2) <sup>c</sup>
NH-2''	7.91, d (6.3)
NH-4''	8.19, d (5.6)

<sup>a-d</sup> The data with the same labels may be interchanged.

**Table S5.**  $^1\text{H}$  NMR (600 MHz), and  $^{13}\text{C}$  NMR (150 MHz) data for berberine (**12**) in  $\text{MeOH-}d_4$ 

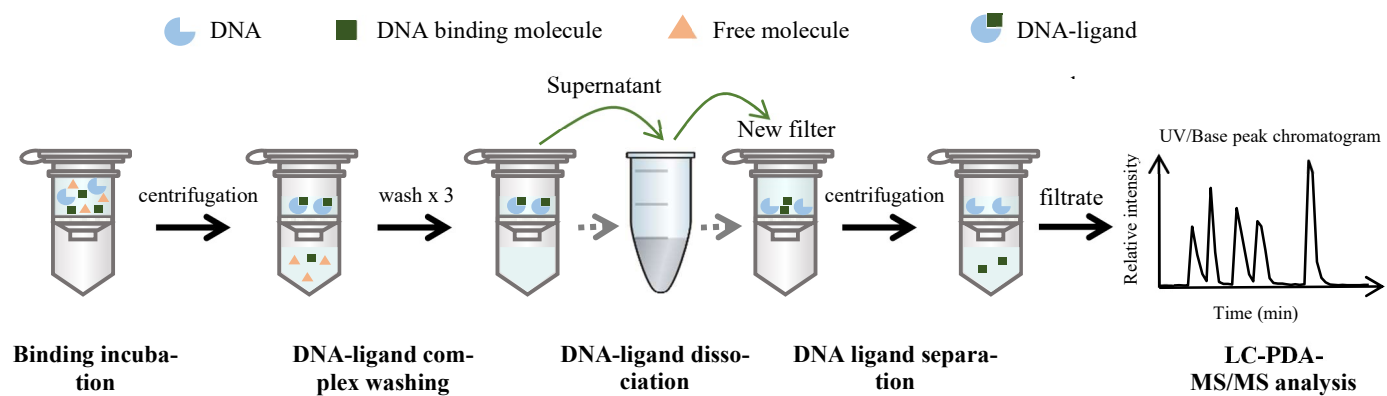
No.	$\delta_{\text{H}}$ ( <i>J</i> in Hz)	$\delta_{\text{C}}$ (ppm)
1	7.64, s	105.3
2		148.7
3		150.8
4	6.94, s	108.2
4a		130.7
5	3.24, t (6.4)	27.0
6	4.91, t (6.4)	56.0
8	9.75, s	145.2
8a		122.1
9		144.6
9-OMe	4.18, s	61.3
10		151.0
10-OMe	4.09, s	56.4
11	7.98, d (8.3)	126.8
12	8.10, d (8.3)	123.3
12a		134.0
13	8.69, s	120.3
13a		138.5
13b		120.7
OCH <sub>2</sub> O	6.09, s	102.5

**Table S6.**  $^1\text{H}$  NMR (600 MHz), and  $^{13}\text{C}$  NMR (150 MHz) data for palmatine (**13**) in  $\text{MeOH-}d_4$ 

No.	$\delta_{\text{H}}$ ( <i>J</i> in Hz)	$\delta_{\text{C}}$ (ppm)
1	7.67, s	109.9
2		150.9
3		153.8
4	7.05, s	112.2
5	3.28, m	27.8
6	4.94, m	56.7
8	9.77, s	146.4
9		145.8
10		151.9
11	8.12, d (9.0)	124.4
12	8.01, d (9.0)	128.1
13	8.81, s	121.3
10-OMe	3.94, s	57.6
12a		135.3
13a		139.8
13b		123.3
2-OMe	3.99, s	57.3
3-OMe	4.11, s	57.0
4a		130.1
8a		120.5
9-OMe	4.21, s	62.5

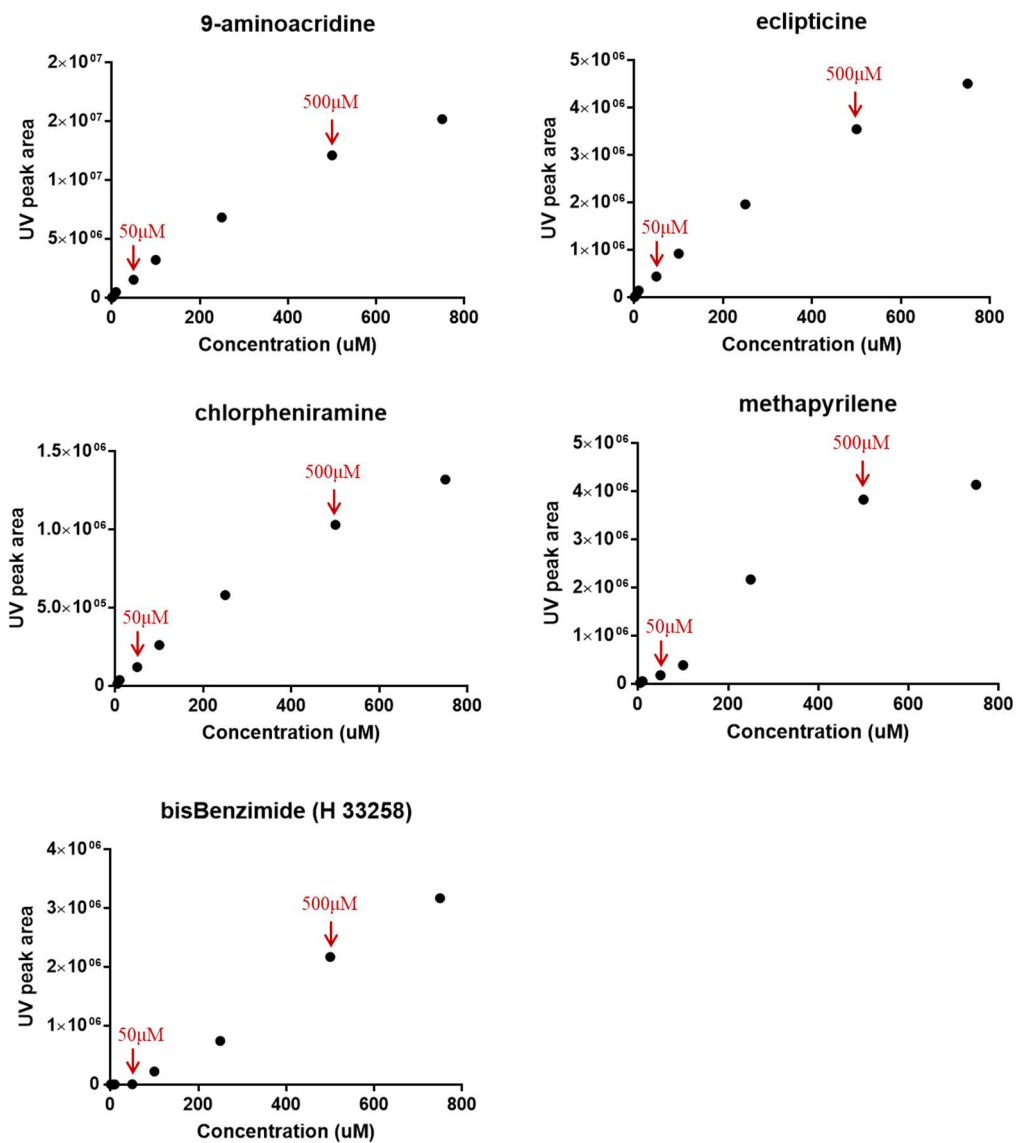
**Table S7.**  $^1\text{H}$  NMR (600 MHz), and  $^{13}\text{C}$  NMR (150 MHz) data for coptisine (**14**) in  $\text{MeOH-}d_4$ 

No.	$\delta_{\text{H}}$ ( $J$ in Hz)	$\delta_{\text{C}}$ (ppm)
1	7.65, s	106.4
2		149.3
3		150.0
4	6.96, s	109.4
5	3.25, m	28.1
6	4.89, m	57.2
8	9.72, s	145.8
9		145.3
10		152.2
11	7.88, d (8.5)	121.9
12	7.86, d (8.5)	123.1
13	8.74, s	122.3
12a		134.4
13a		139.0
13b		122.5
4a		131.8
8a		113.7
$\text{OCH}_2\text{O}$	6.47, s	106.2
$\text{OCH}_2\text{O}$	6.11, s	103.7

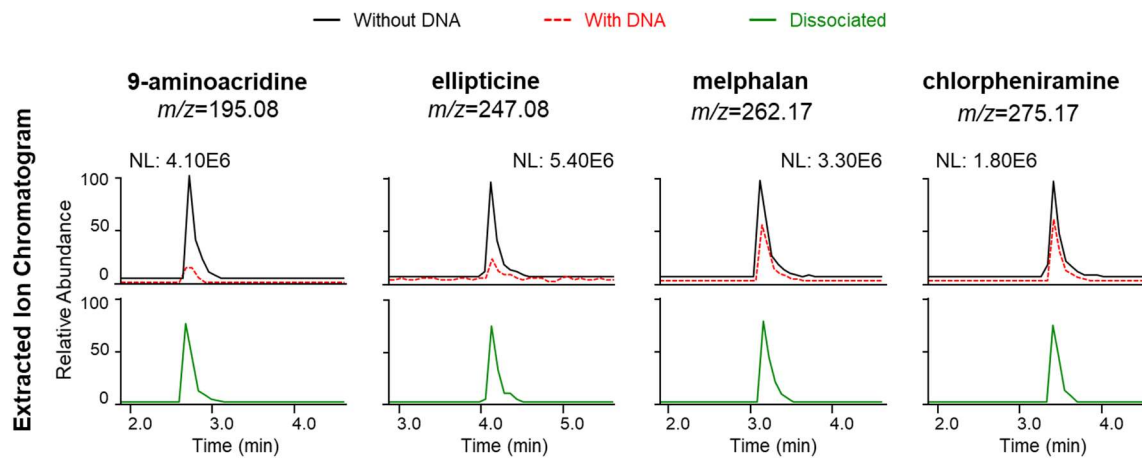


**Figure S1.** Overflow of DNA-ligand complex dissociation

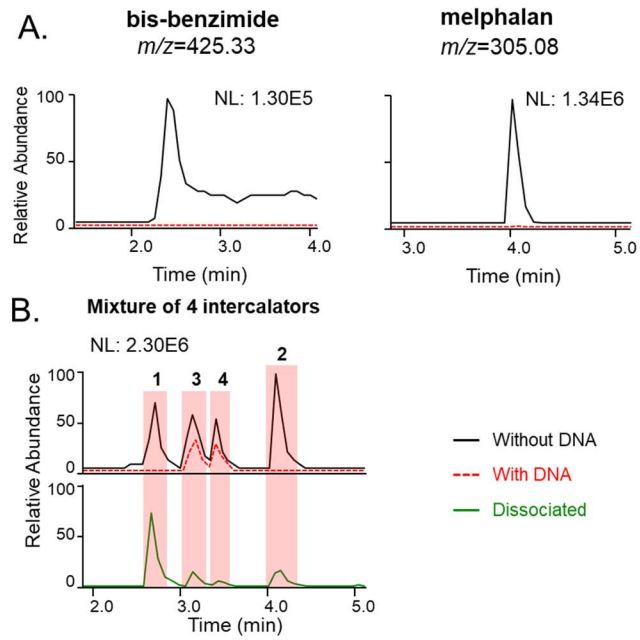




**Figure S2.** Calibration curve analysis (UV peak area) of 4 intercalators and 1 groove binder

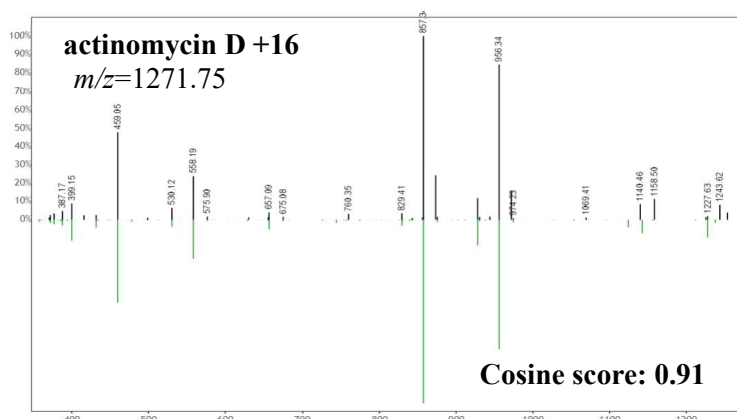
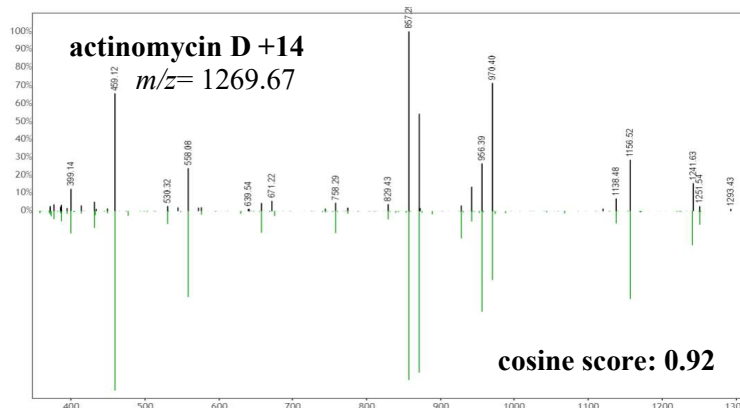
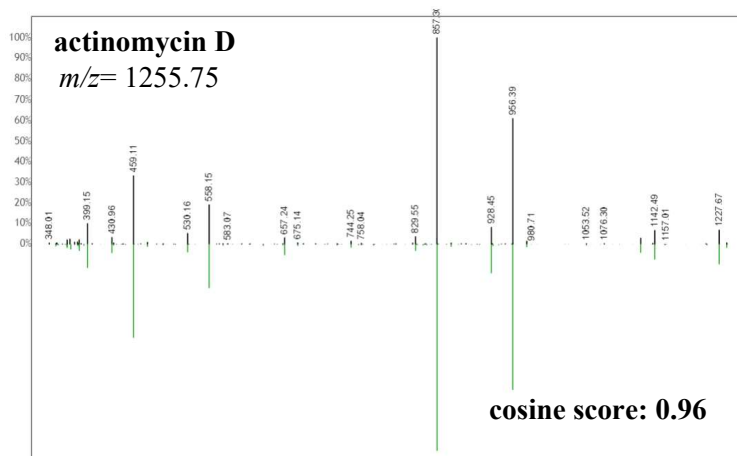


**Figure S3.** EIC analysis of the 4 individual intercalators in the assay.

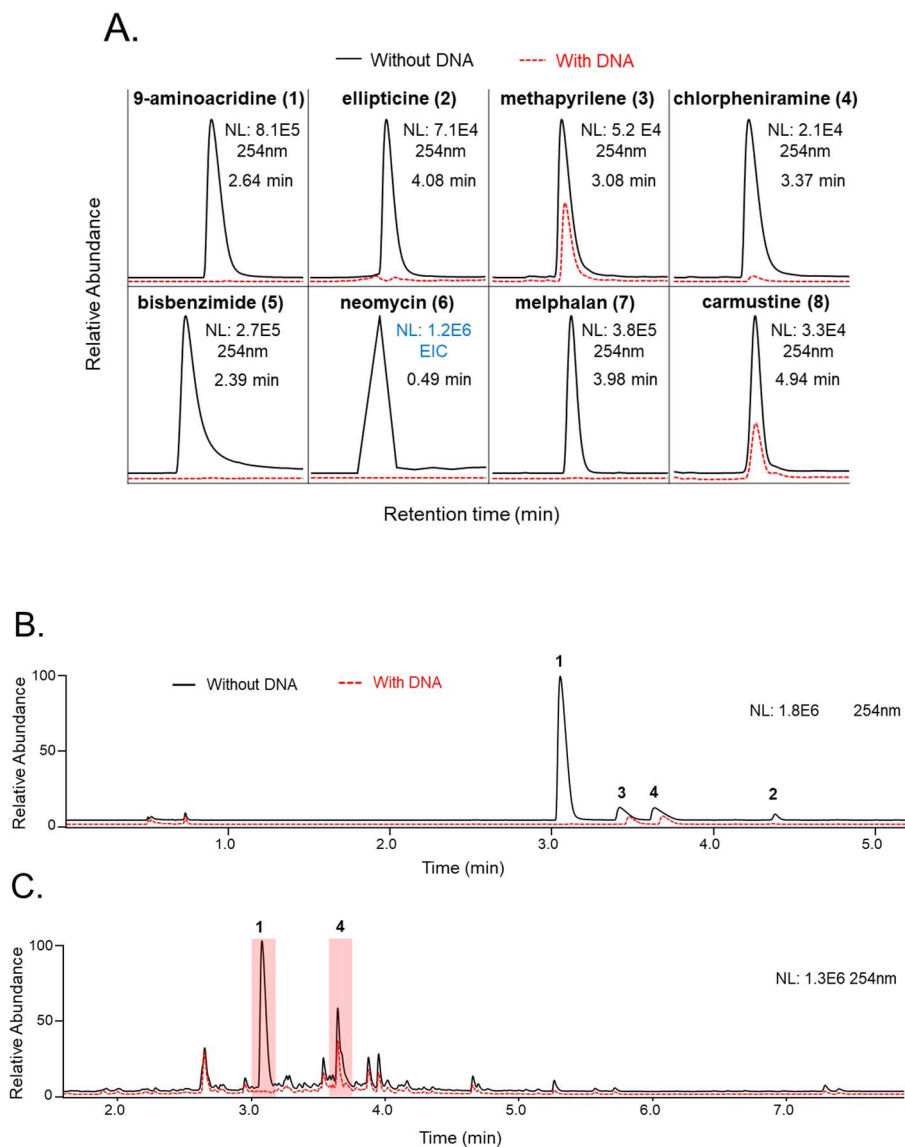


**Figure S4.** EIC analysis of bisbenzimidazole (H33258), melphalan, and 4 intercalators mixture.

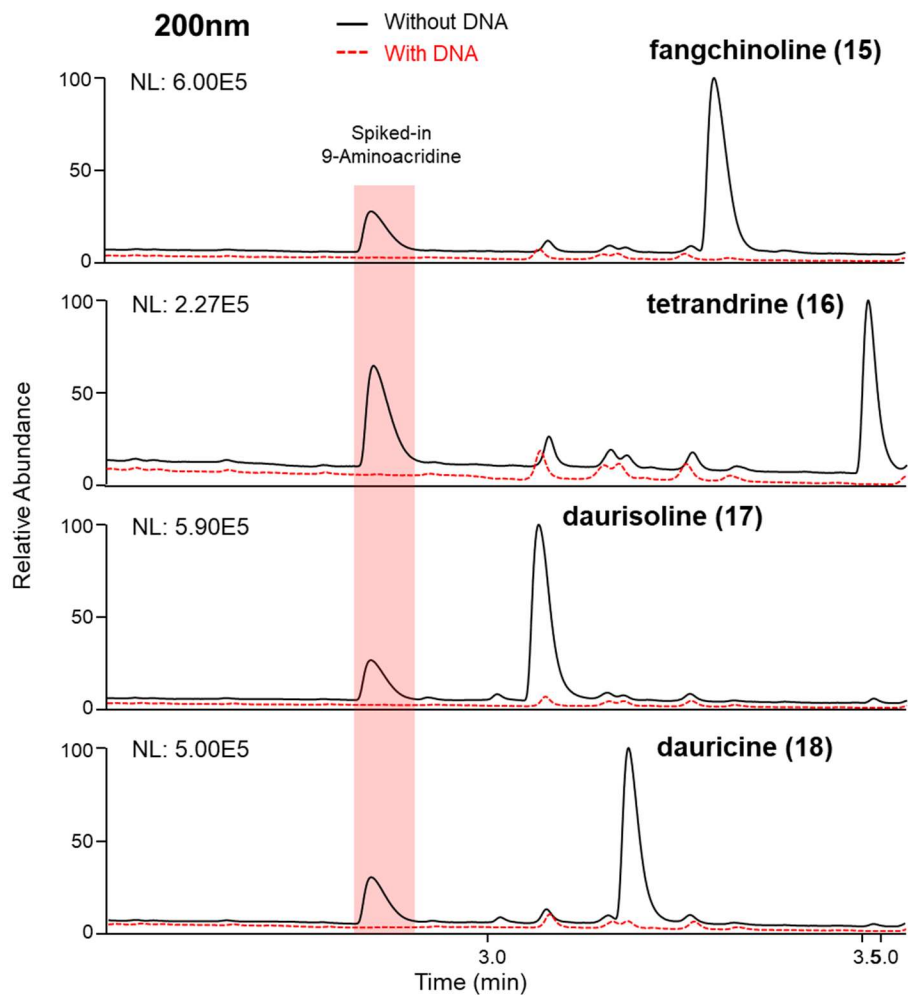
<b><i>m/z</i></b>	1255.75	1269.67	1271.75
<b>RT (min)</b>	8.64	8.64	7.77
<b>annotation</b>	actinomycin D	actinomycin D +14 Da	actinomycin D +16 Da
<b>cosine score</b>	0.96	0.92	0.91
<b><i>m/z</i> error (Da)</b>	0.11	0.1	0.12
<b>reference Spectrum ID</b>	CCMSLIB00000006871	CCMSLIB000000081798	CCMSLIB00000006871



**Figure S5.** Annotation and mirror plots of actinomycin D and its analogues using GNPS.

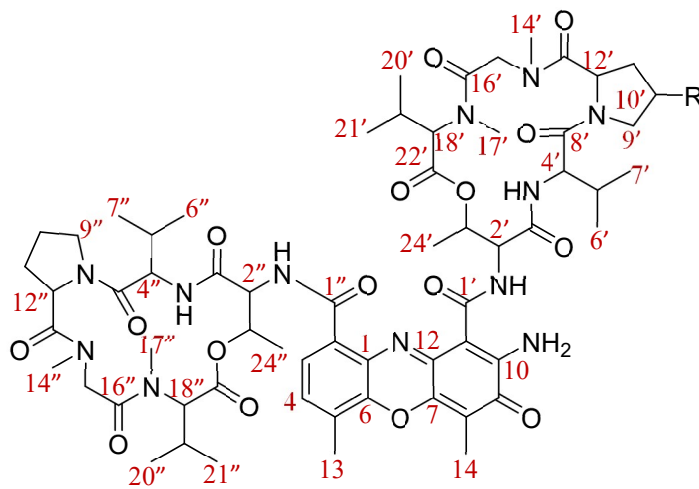
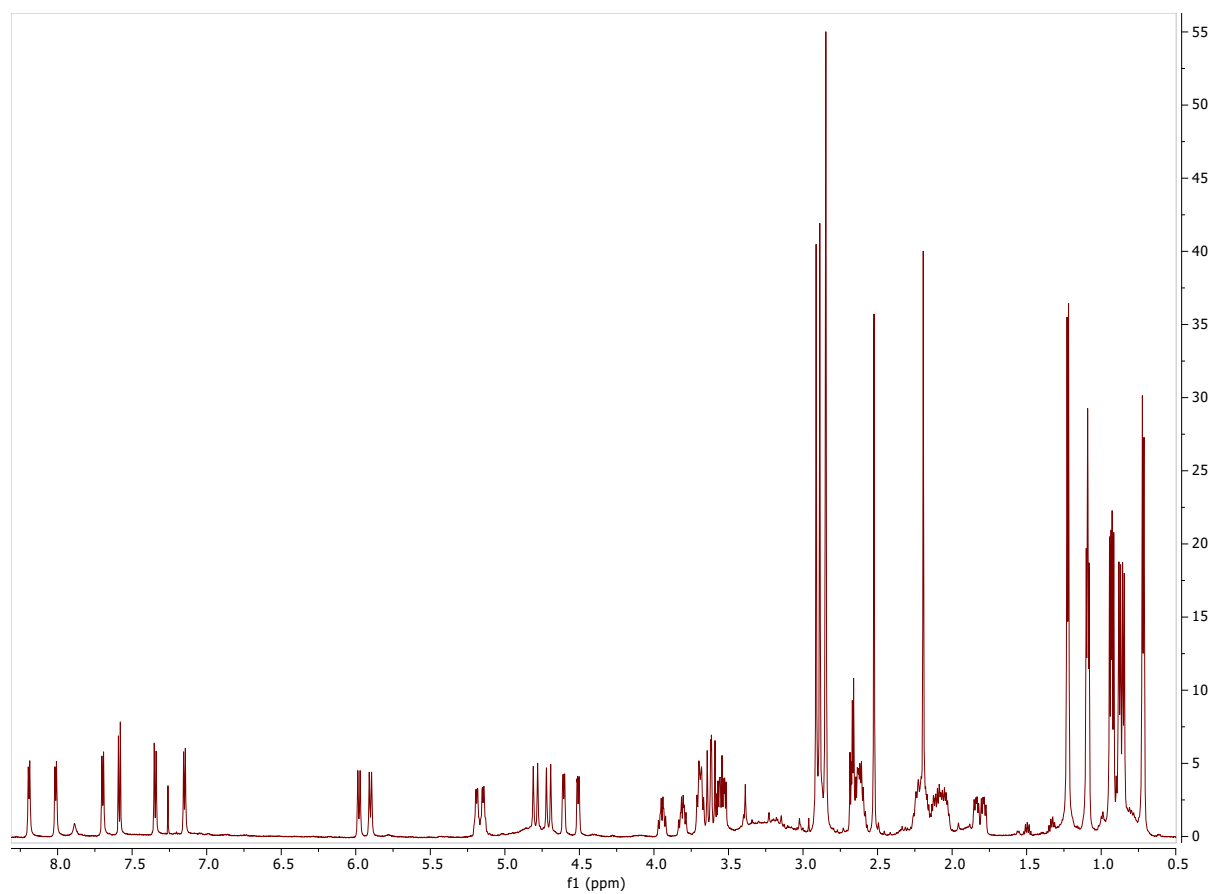


**Figure S6.** Results of the high throughput LLAMAS development. **A)** Performance of the 8 known binders: 4 intercalators, including: 9-aminoacridine (34  $\mu$ M), ellipticine (54  $\mu$ M), methapyrilene (45  $\mu$ M), and chlorpheniramine (34  $\mu$ M); 2 groove binders: bisBenzimidazole (H 33258) (125g  $\mu$ M), and neomycin (157  $\mu$ M), and 2 covalent binders: melphalan (44  $\mu$ M) and carmustine (62  $\mu$ M) in the assay. Activity of neomycin was confirmed by analyzing EIC peaks extracted at m/z of 615.09-615.51; while others were confirmed by UV-Trace analysis. (Here, the concentrations are all final ones in binding incubation step) LC separation gradient: 7-50% B. **B)** Performance of the mixture of 4 intercalators in the assay: **1:** 9-aminoacridine, **2:** methapyrilene, **3:** chlorpheniramine, **4:** ellipticine. **C)** Known DNA intercalators were successfully characterized in plant extract spiked with 9-aminoacridine (**1**) and ellipticine (**4**) with the ratio of 1:5:250 (w:w:w).



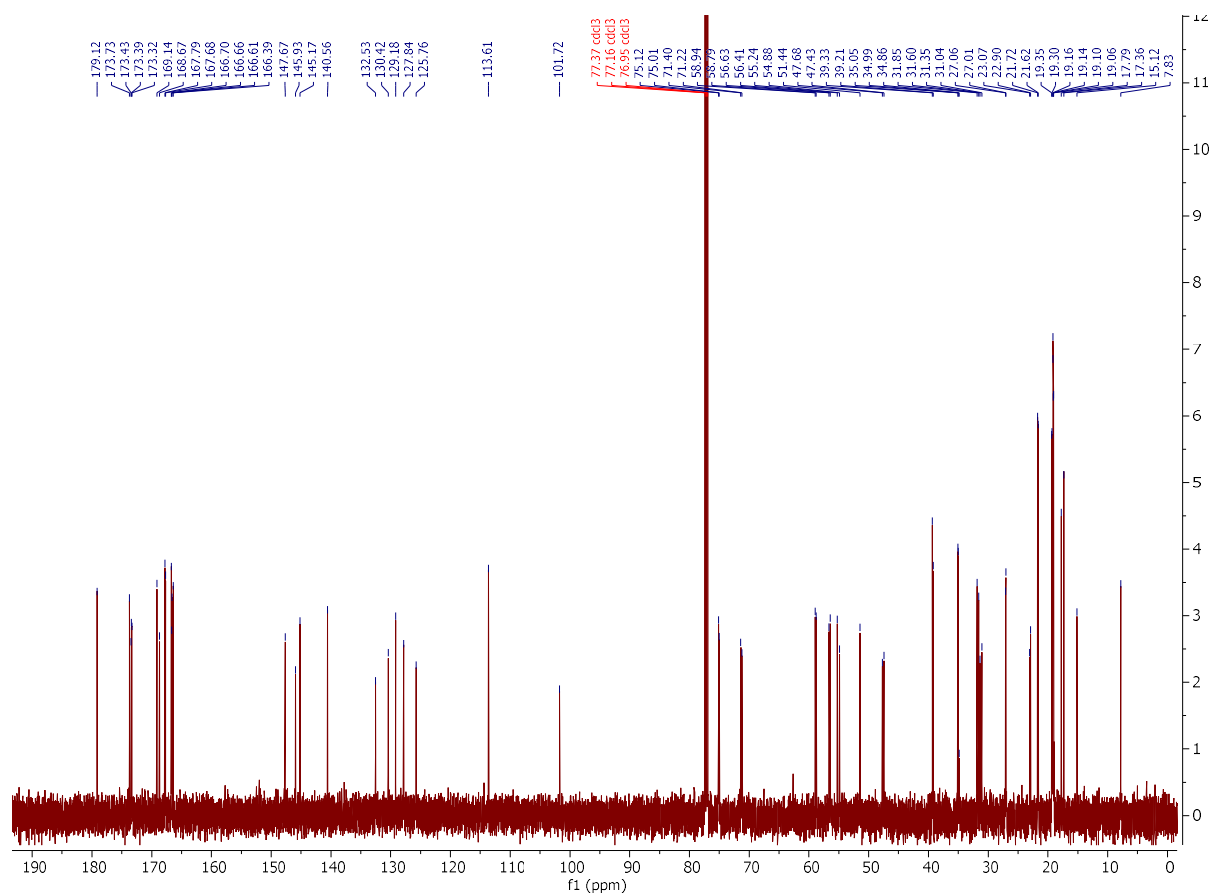
**Figure S7.** Confirmation of the DNA binding activities of fangchinoline (15), tetrandrine (16), daurisoline (17), and dauricine (18). UV chromatogram at 200 nm showed their peaks in experimental group (red line) disappeared, while detected in control group, indicating their DNA binding activities.

**Figure S8.**  $^1\text{H}$  NMR (600 MHz) spectrum of actinomycin D (**9**) in  $\text{CDCl}_3$



Actinomycin D: R=H  
Actinomycin V: R=O  
Actinomycin X<sub>0β</sub>: R=OH

**Figure S9.**  $^{13}\text{C}$  NMR (600 MHz) spectrum of actinomycin D (**9**) in  $\text{CDCl}_3$





**Figure S10.**  $^1\text{H}$  NMR (600 MHz) spectrum of actinomycin V (**10**) in  $\text{CDCl}_3$

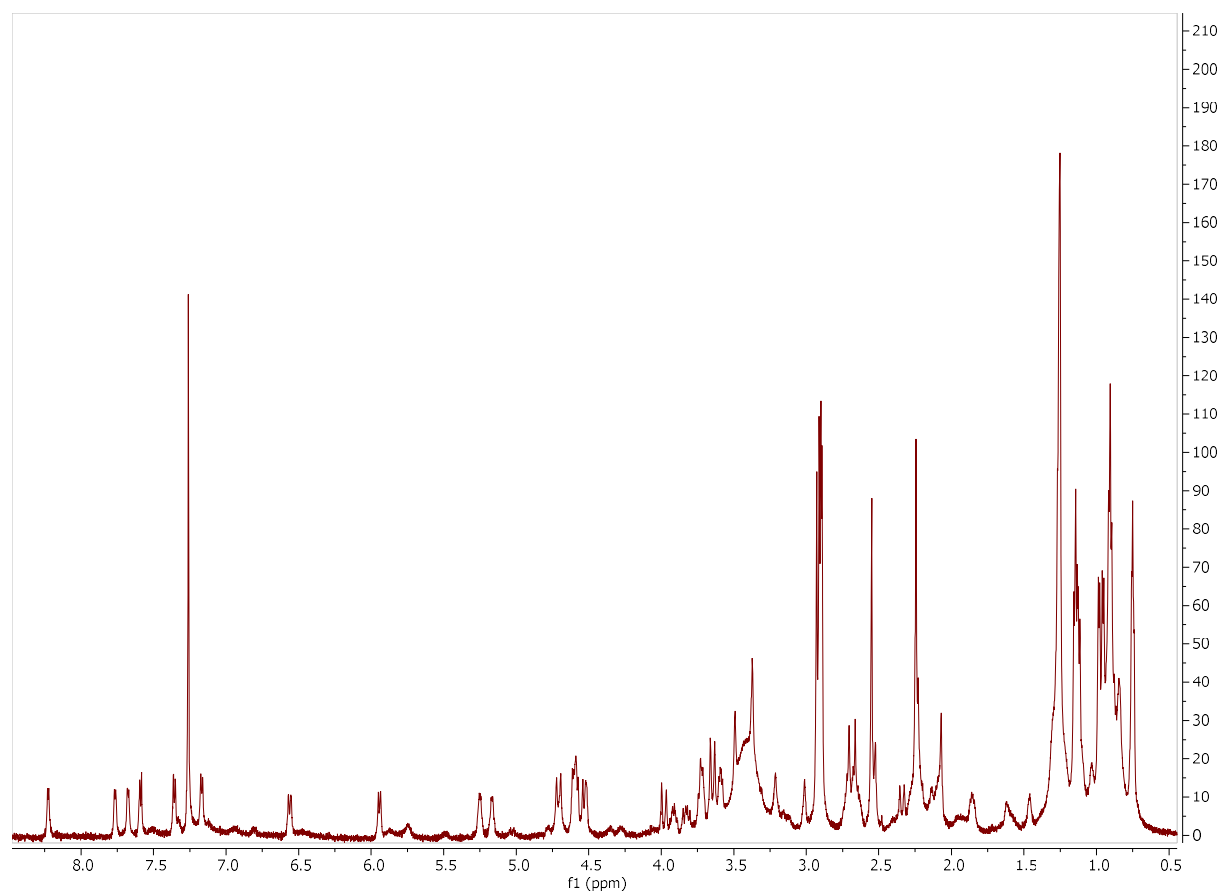
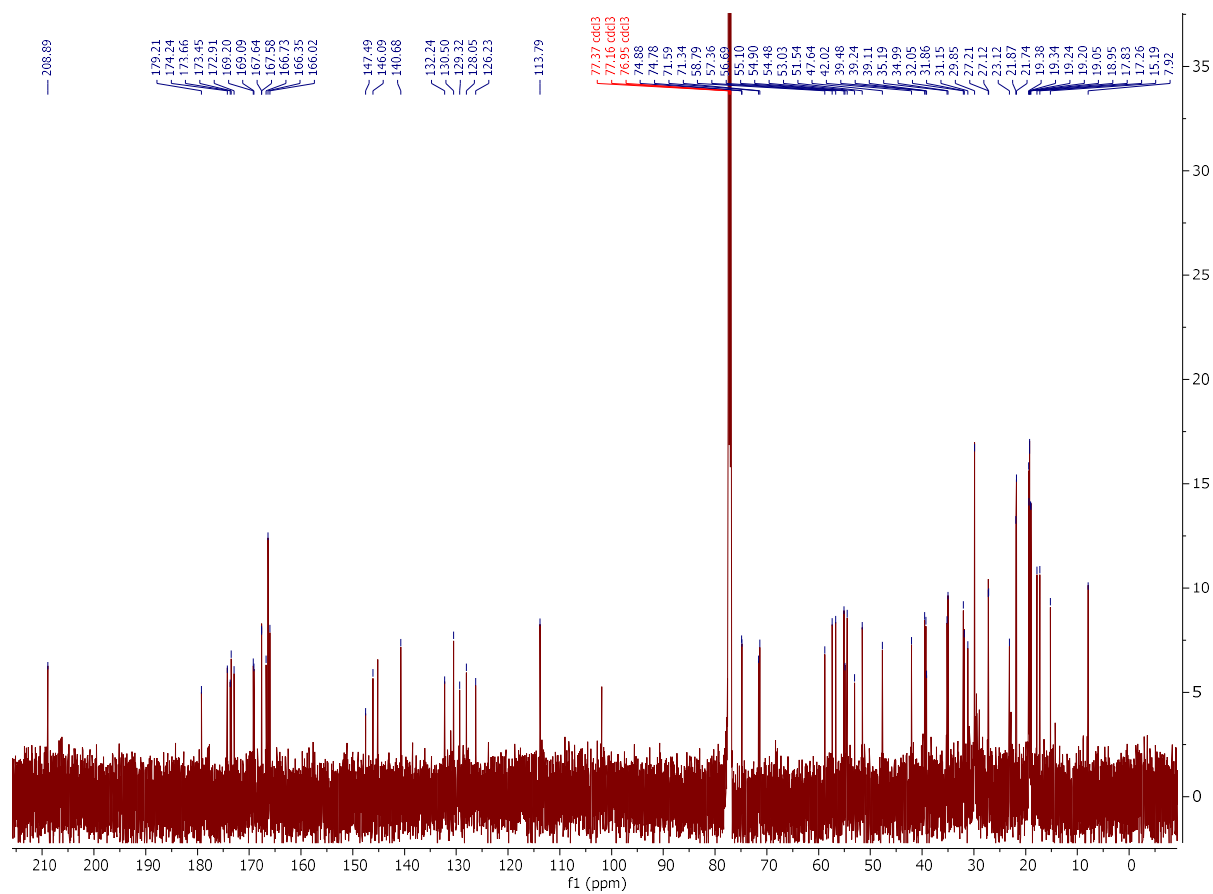
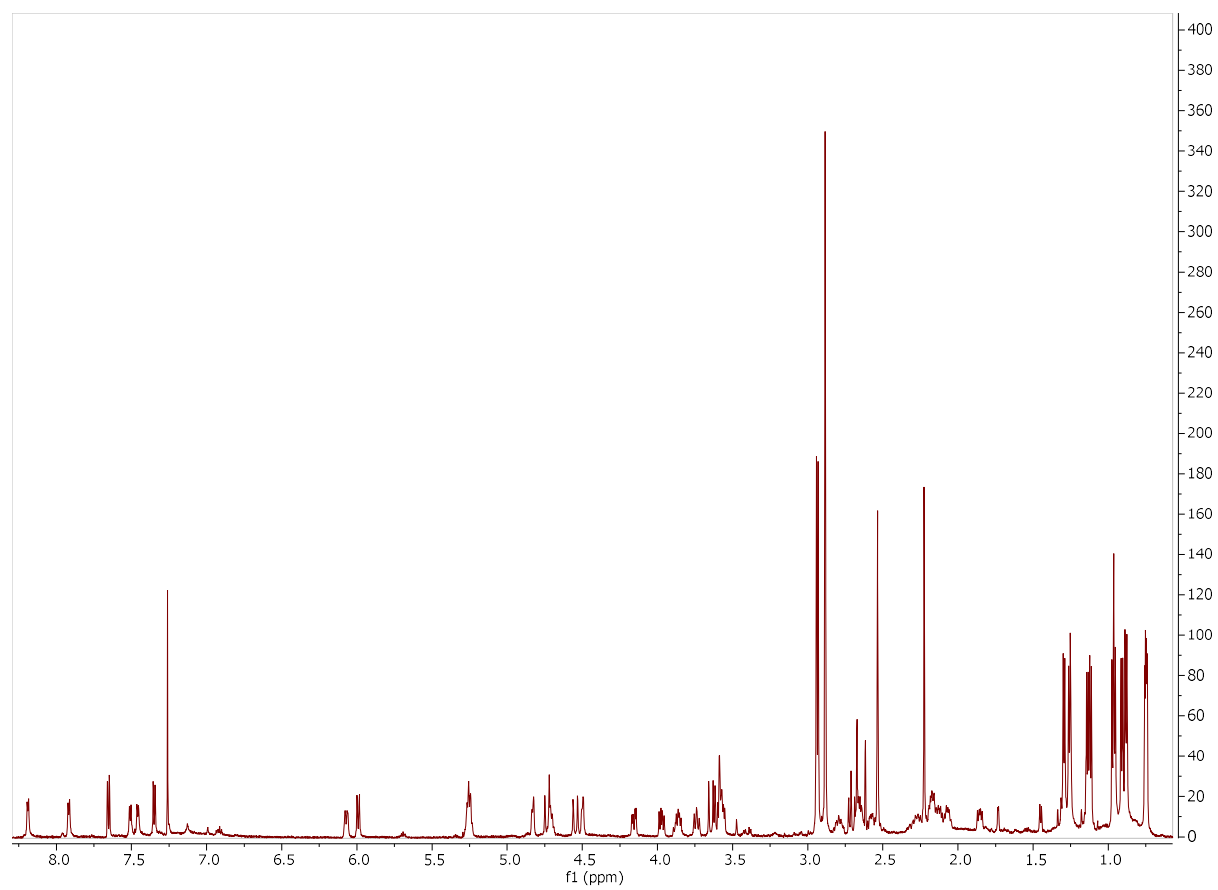


Figure S11.  $^{13}\text{C}$  NMR (600 MHz) spectrum of actinomycin V (10) in  $\text{CDCl}_3$



**Figure S12.**  $^1\text{H}$  NMR (600 MHz) spectrum of actinomycin  $X_{0\beta}$  (**11**) in  $\text{CDCl}_3$



**Figure S13.**  $^1\text{H}$  NMR (600 MHz) spectrum of berberine (**12**) in  $\text{MeOH-}d_4$

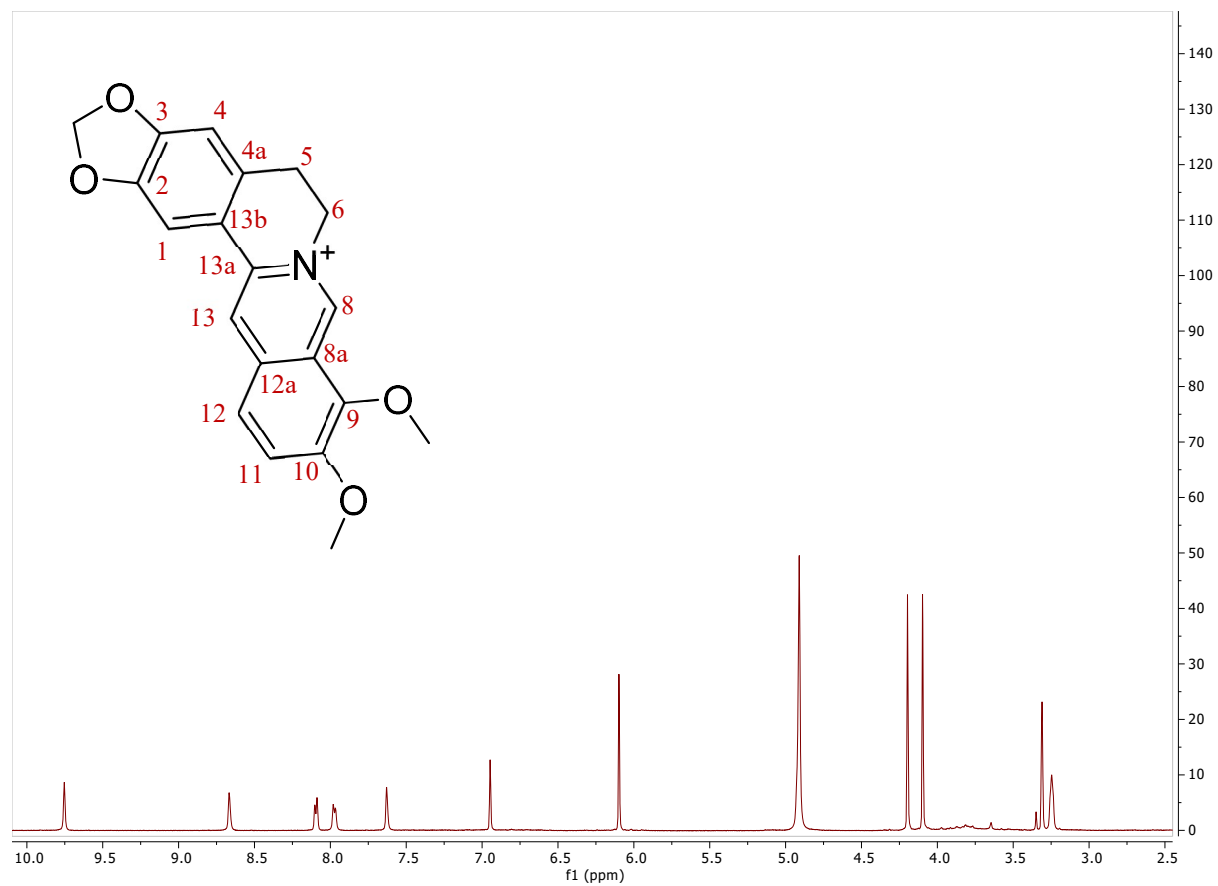
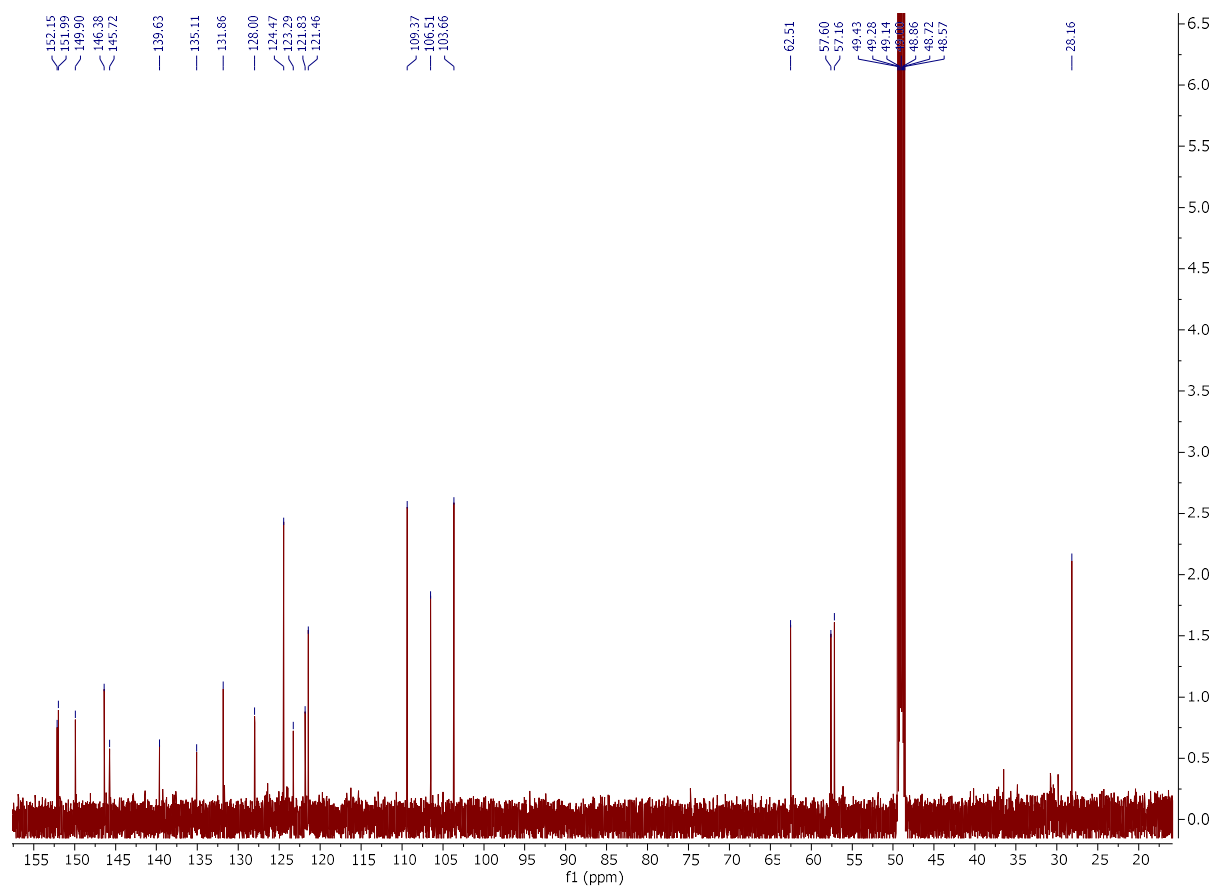
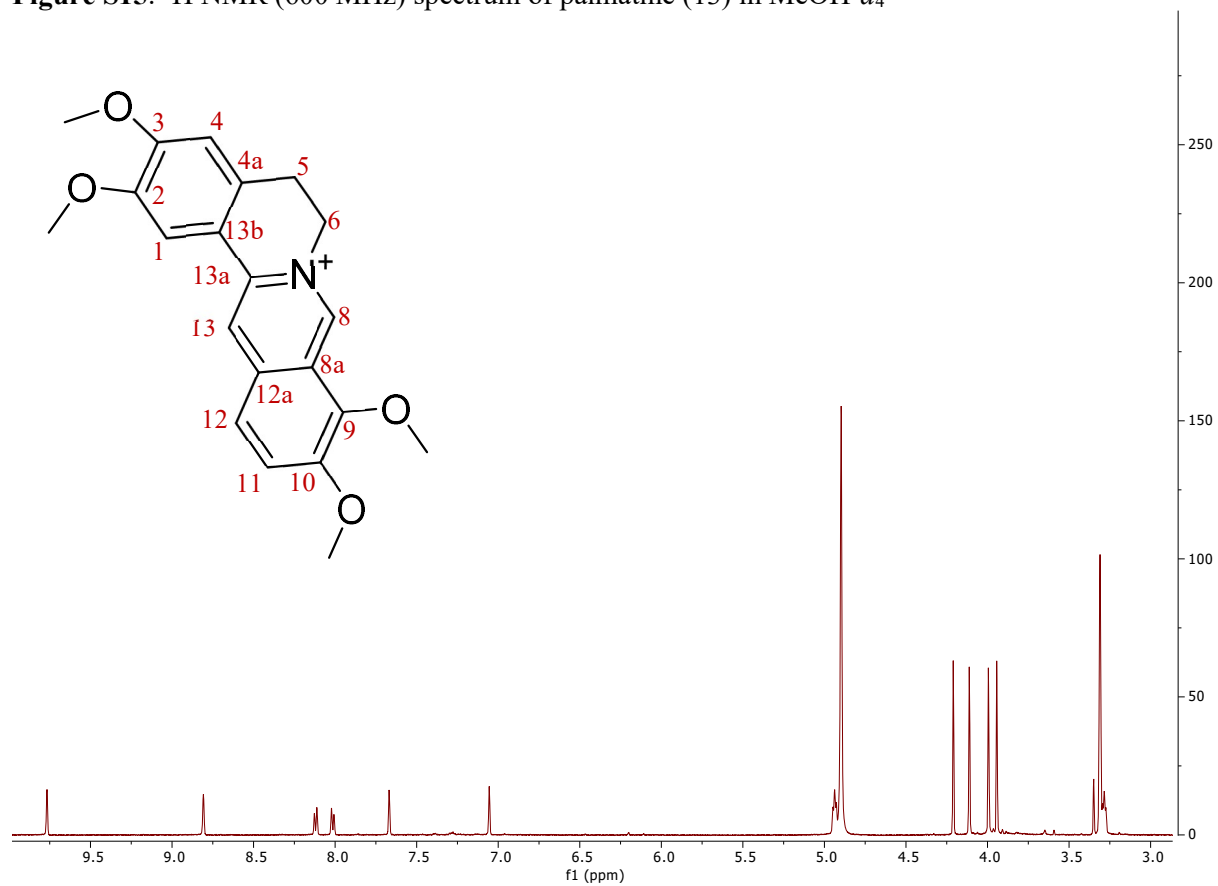


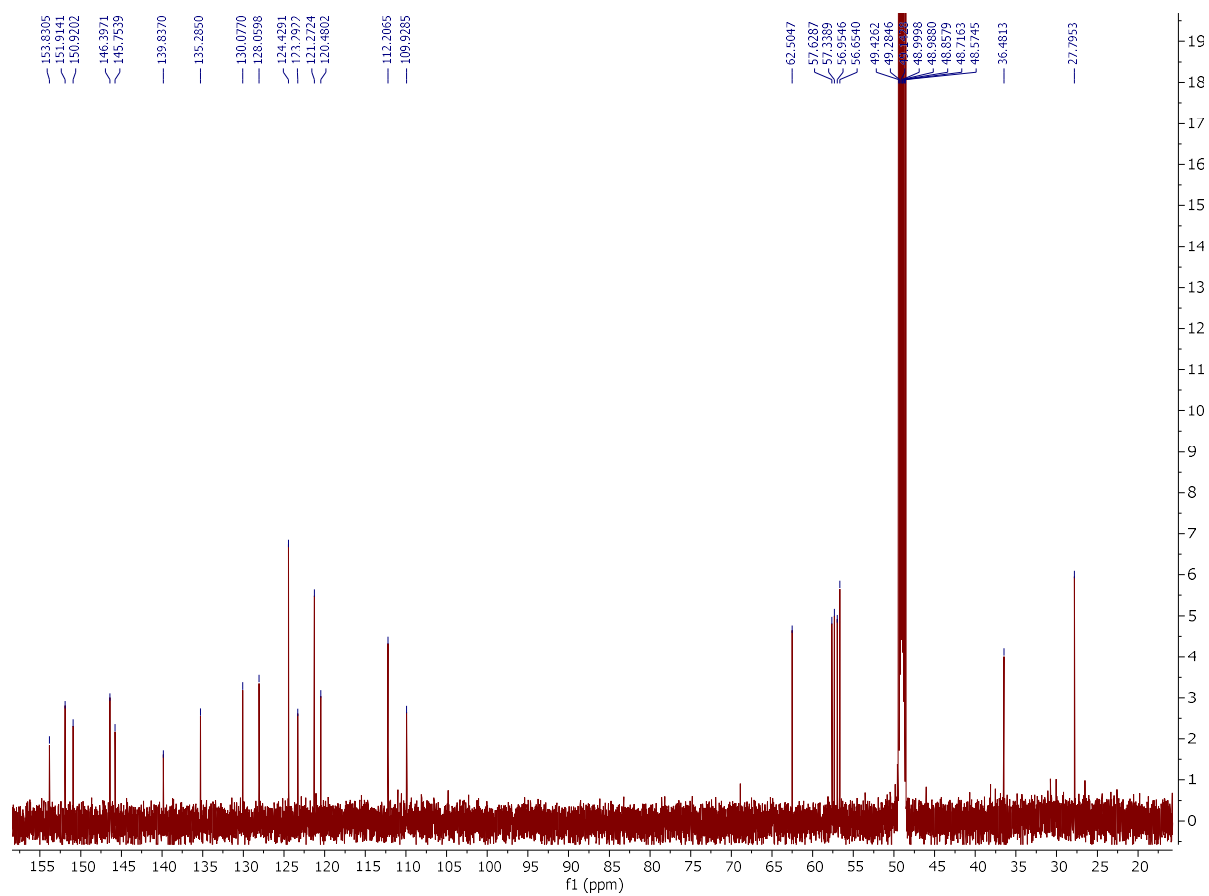
Figure S14.  $^{13}\text{C}$  NMR (600 MHz) spectrum of berberine (**12**) in  $\text{MeOH-}d_4$



**Figure S15.**  $^1\text{H}$  NMR (600 MHz) spectrum of palmatine (13) in  $\text{MeOH-}d_4$



**Figure S16.**  $^{13}\text{C}$  NMR (600 MHz) spectrum of palmatine (**13**) in  $\text{MeOH-}d_4$



**Figure S17.**  $^1\text{H}$  NMR (600 MHz) spectrum of coptisine (**14**) in  $\text{MeOH-}d_4$

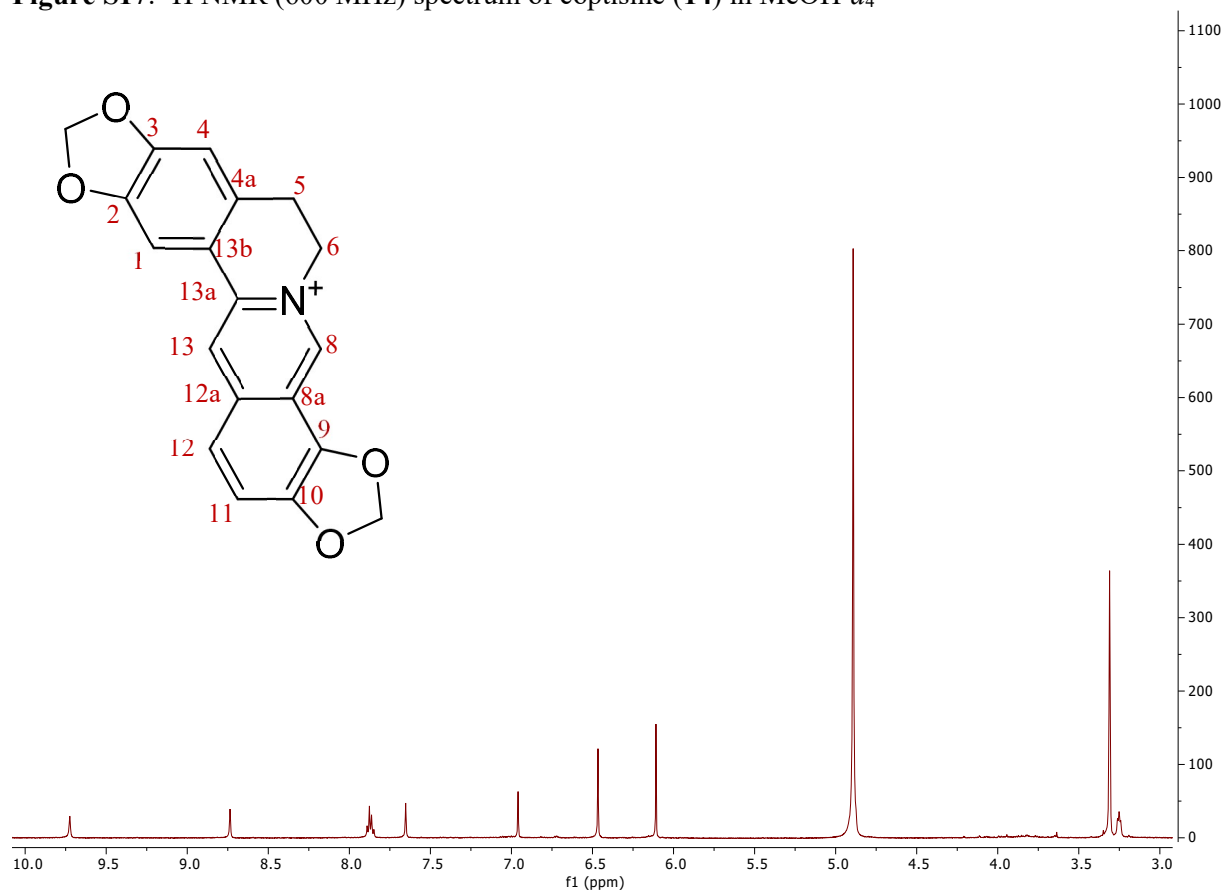




Figure S18.  $^{13}\text{C}$  NMR (600 MHz) spectrum of coptisine (**14**) in  $\text{MeOH-}d_4$

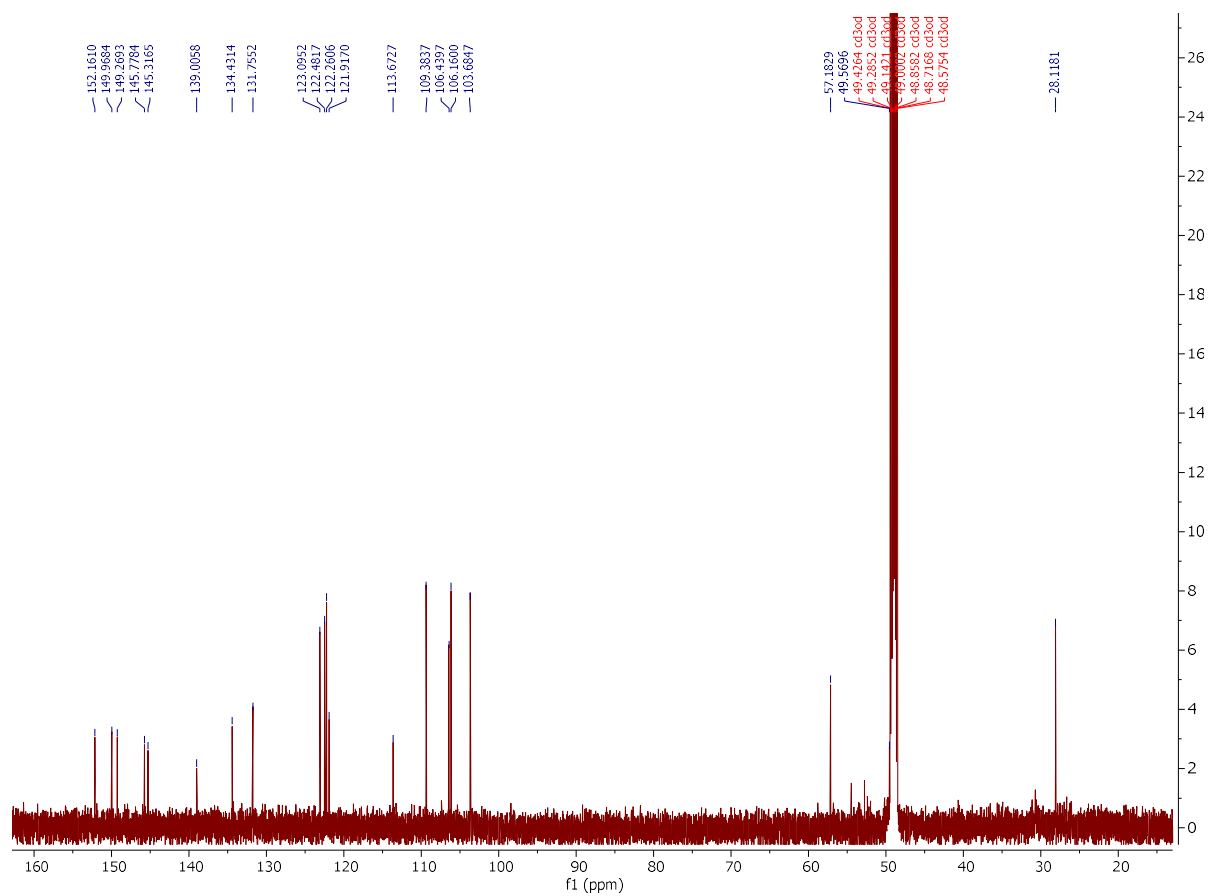


Figure S19.  $^1\text{H}$  NMR (500 MHz) spectrum of fangchinoline (**15**) in  $\text{CDCl}_3$

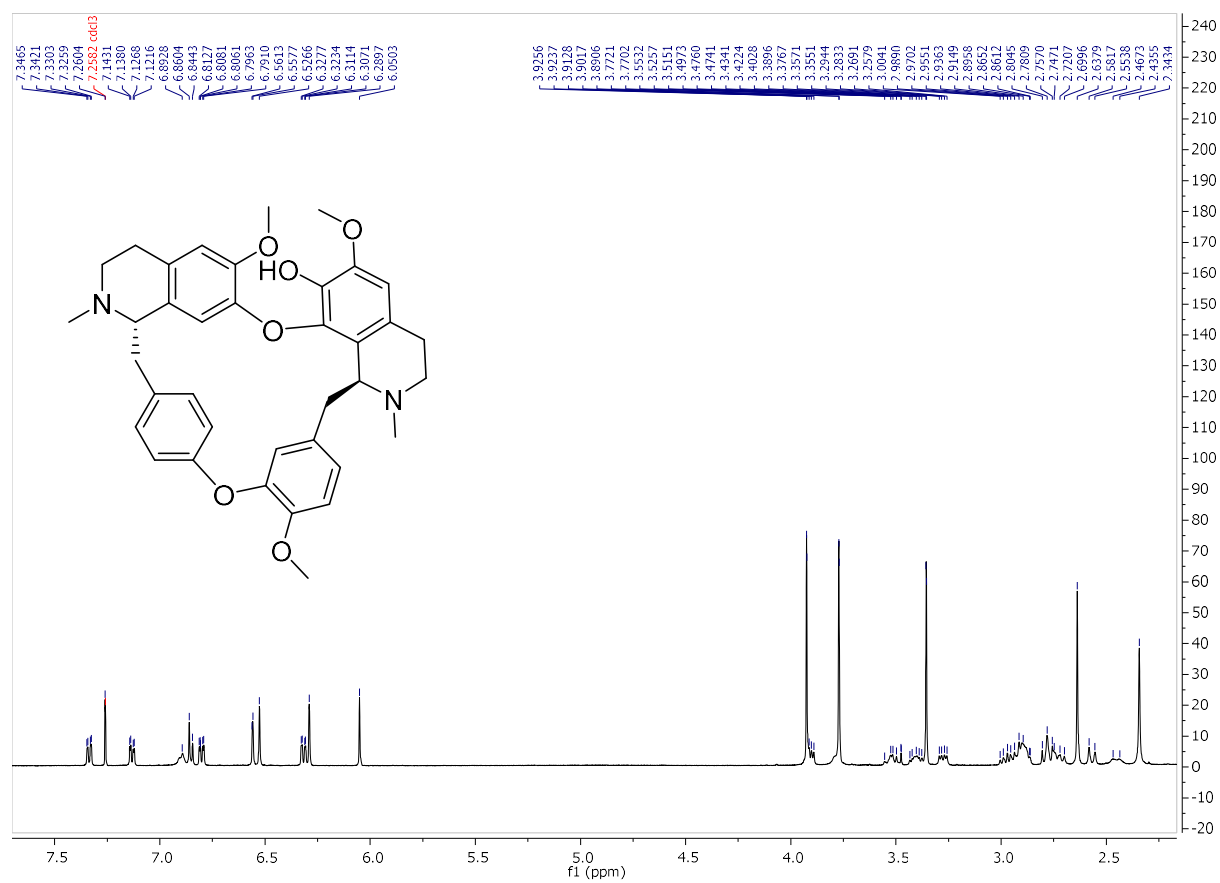


Figure S20. <sup>1</sup>H NMR (500 MHz) spectrum of tetrandrine (16) in CDCl<sub>3</sub>

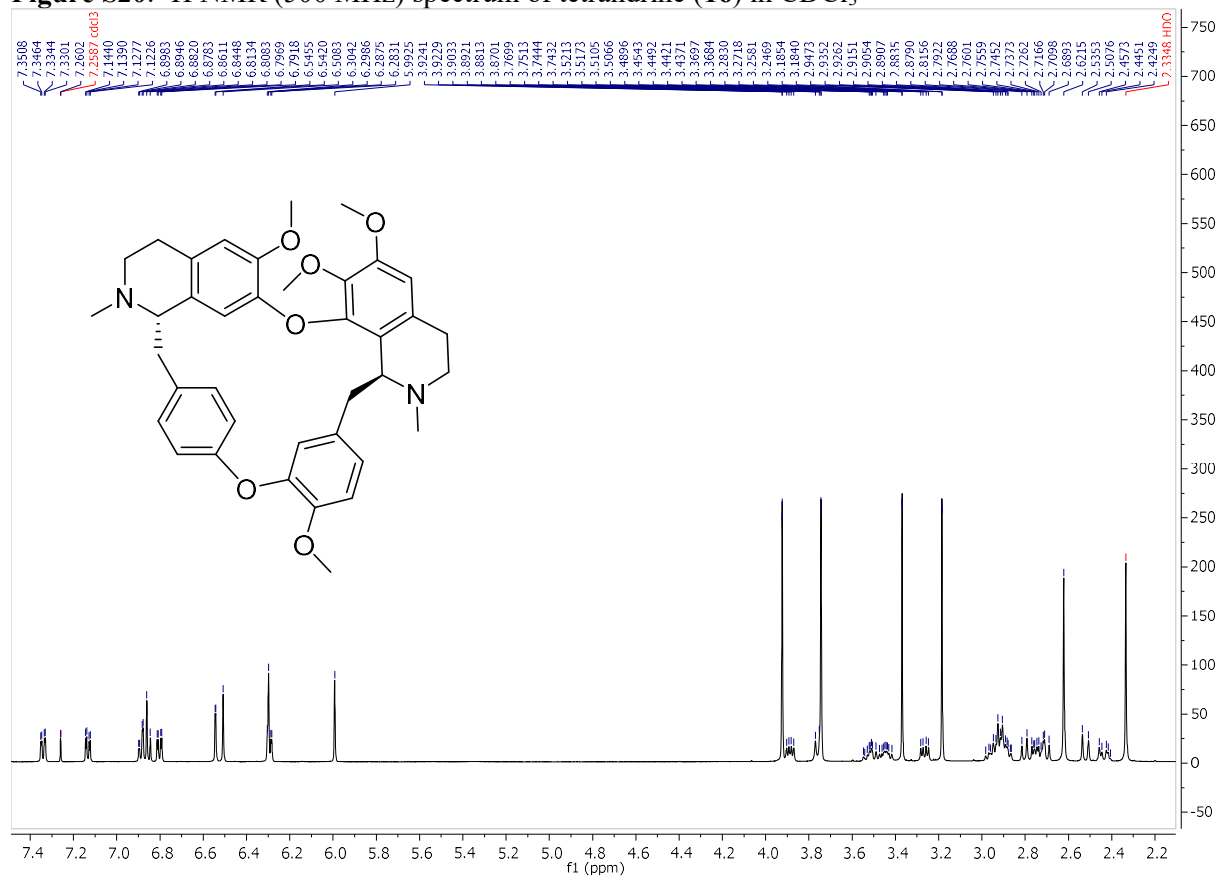


Figure S21. <sup>1</sup>H NMR (500 MHz) spectrum of daurisoline (17) in CDCl<sub>3</sub>

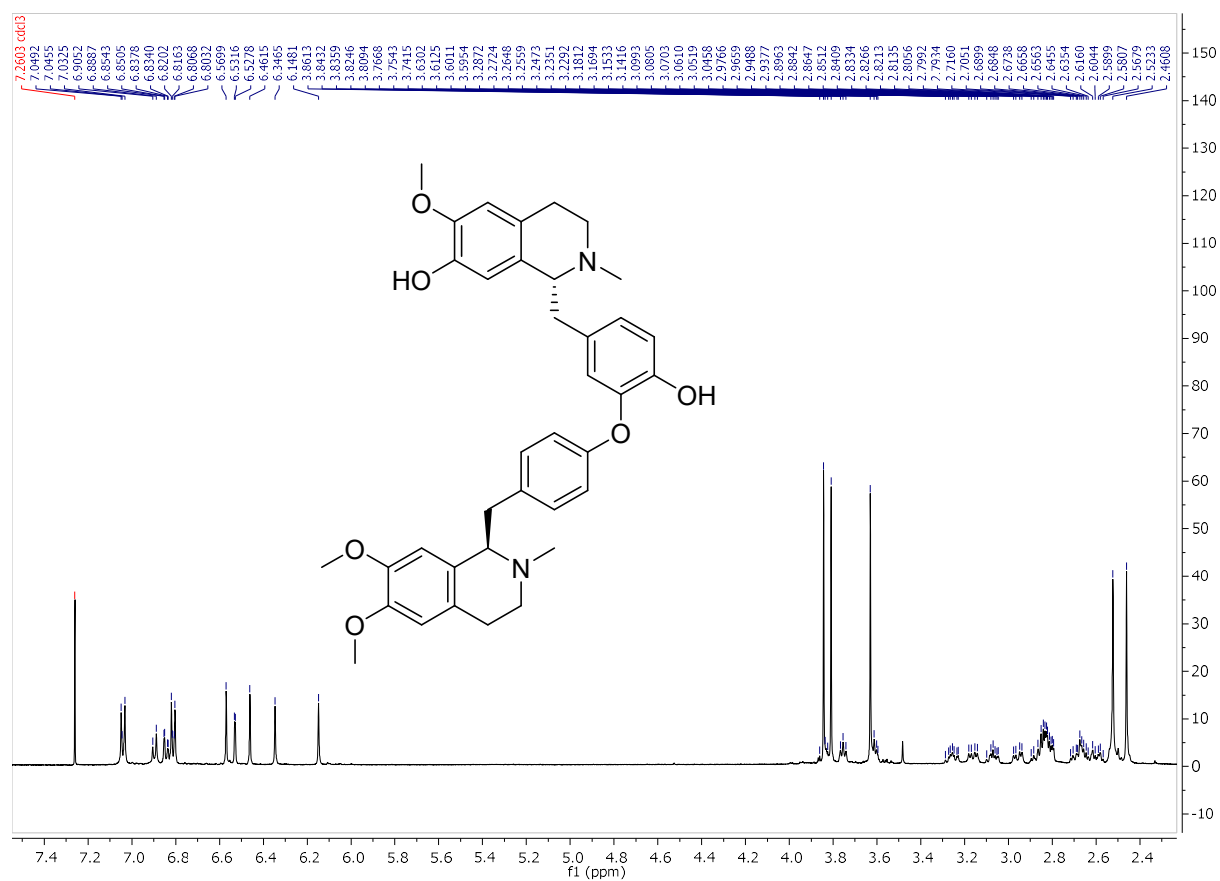
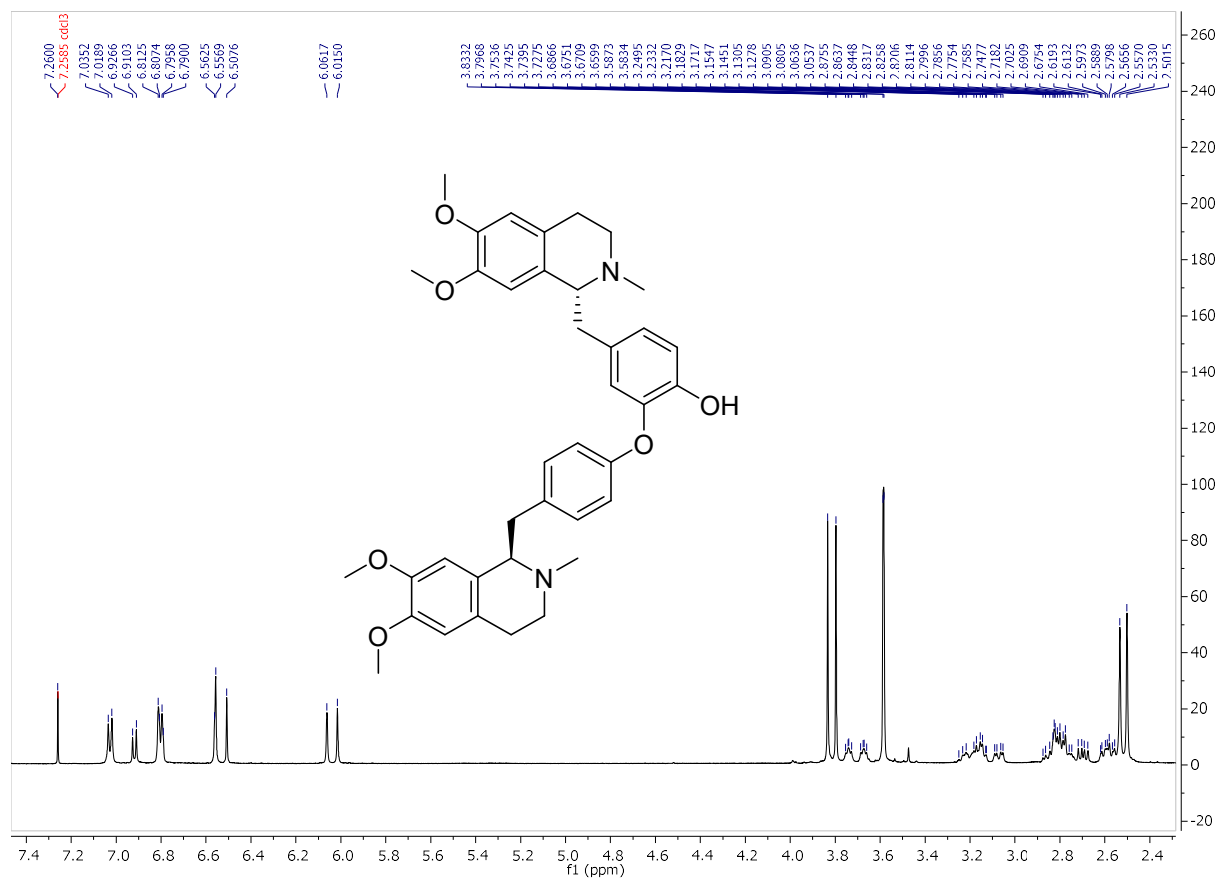


Figure S22. <sup>1</sup>H NMR (500 MHz) spectrum of dauricine (18) in CDCl<sub>3</sub>



## Supporting information for chapter 4

6/20/2019

RightsLink - Your Account

<b>SPRINGER NATURE LICENSE TERMS AND CONDITIONS</b>	
Jun 20, 2019	
<p>This Agreement between Hongyan Ma ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.</p>	
<b>License Number</b>	4613250869205
<b>License date</b>	Jun 20, 2019
<b>Licensed Content Publisher</b>	Springer Nature
<b>Licensed Content Publication</b>	Journal of The American Society for Mass Spectrometry
<b>Licensed Content Title</b>	Finding Biomass Degrading Enzymes Through an Activity-Correlated Quantitative Proteomics Platform (ACPP)
<b>Licensed Content Author</b>	Hongyan Ma, Daniel G. Delafield, Zhe Wang et al
<b>Licensed Content Date</b>	Jan 1, 2017
<b>Licensed Content Volume</b>	28
<b>Licensed Content Issue</b>	4
<b>Type of Use</b>	Thesis/Dissertation
<b>Requestor type</b>	non-commercial (non-profit)
<b>Format</b>	print and electronic
<b>Portion</b>	full article/chapter
<b>Will you be translating?</b>	no
<b>Circulation/distribution</b>	>50,000
<b>Author of this Springer Nature content</b>	yes
<b>Title</b>	BIOACTIVE MOLECULES DETECTION AND IDENTIFICATION IN COMPLEX BIOLOGICAL SAMPLES THROUGH BIOASSAY-GUIDED QUANTITATIVE METABOLOMICS AND PROTEOMICS STUDIES
<b>Institution name</b>	University of Oklahoma
<b>Expected presentation date</b>	May 2020
<b>Order reference number</b>	4613250360139
<b>Requestor Location</b>	2920 Chautauqua Ave. Apt7 2920 Chautauqua Ave. Apt7  NORMAN, OK 73072 United States Attn: Hongyan Ma
<b>Total</b>	0.00 USD
<b>Terms and Conditions</b>	
<b>Springer Nature Customer Service Centre GmbH Terms and Conditions</b>	
This agreement sets out the terms and conditions of the licence (the Licence) between you and Springer Nature Customer Service Centre GmbH (the Licensor). By clicking 'accept' and completing the transaction for the material (Licensed Material), you	

<https://s100.copyright.com/MyAccount/web.jsp/viewprintablelicensefrommyorders.jsp?ref=1f0bf53e-4499-4e4a-b9e1-c77be7886acb>

1/4



RightsLink®



Home



Help



Email Support



Hongyan Ma ▾

**Finding Biomass Degrading Enzymes Through an Activity-Correlated Quantitative Proteomics Platform (ACPP)**

Author: Hongyan Ma, Daniel G. Delafield, Zhe Wang, et al

Publication: Journal of the American Society for Mass Spectrometry

Publisher: American Chemical Society

Date: Apr 1, 2017

Copyright © 2017, American Chemical Society

**PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE**

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

[BACK](#)[CLOSE WINDOW](#)