IDENTIFICATION AND CHEMICAL CHARACTERIZATION OF MARINE NATURAL PRODUCTS USING UPLC-QTOF-MS COUPLED TO A NOVEL INFORMATICS PLATFORM

Roger G. Linington[1], Kenji L. Kurita[1], Jimmy Yuk[2], Kate Yu[2], Mark Wrona[2] and Giorgis Isaac[2]

[1]Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada; [2]Waters Corporation, 34 Maple Street, Milford, MA 01757

HIGHLIGHTS

- Developed a high-throughput and automated identification of marine microbial compounds using high-resolution UPLC/QTof MS technology and Natural Product Application Solution (NPAS) with UNIFI.
- Provide a comprehensive annotation of the identities and biological attributes of all bioactive constituents.

INTRODUCTION

Despite the widespread use of natural products as inspirations for existing drugs, the rate of development of new natural product-based drugs has slowed in recent years. This is due in part to an increase in the rates of rediscovery of existing scaffolds, as well as limited early characterization of biological functions of hits from primary screens.

Despite the advances in hardware and the development of numerous derivatization, labeling, and analytical methods for compound identification, detailed and unequivocal determination of the constitution and configuration of complex natural product is a time-consuming task that requires a significant investment of resources and materials. Bioinformatics tools are becoming essential in natural products research, as advances in experimental throughput and the complexity of data obtained from biological profiling make manual interpretation difficult or impossible (1).

The Waters Natural Products Application Solution (NPAS) with UNIFI provides a novel and comprehensive strategy for NP data analysis (2). It utilizes the ACQUITY UPLC® I-Class and Xevo® G2-XS QTof MS to acquire MS^E data (data-independent acquisition that provides exact mass precursor and corresponding fragment ions for identification and structural elucidation), which are then searched against custom library integrated within UNIFI software. The structures of the matched components are automatically verified by MassFragment™ with corresponding fragment ions.

We have developed a new platform for integrated natural products discovery that uses image-based screening and high-resolution UPLC/QTof to permit 'function-first' annotation of natural products libraries and provide a comprehensive annotation of the identities and biological attributes of all bioactive constituents. The image-based screening and high resolution mass spectrometry work flow is shown in figure 1.

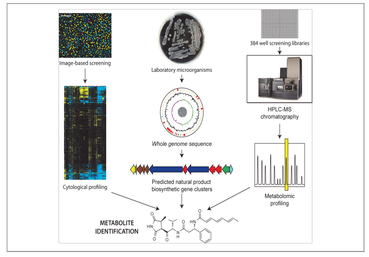

Figure 1. High resolution mass spectrometry and image-based screening work flow platform for compound discovery from complex natural products mixtures.

METHODS

Sample Preparation

Microbes were isolated from marine sediment off Panama, grown under standard fermentation conditions with XAD-16 resin, extracted with 1:1 methanol/dichloromethane, and fractionated on a reverse phase C_{18} column with an eleutropic series of water and methanol (20%, 40%, 60%, 80%, 100% methanol, and EtOAc) after first washing off polar molecules with 10% methanol in water. These fractionated extracts or prefractions were dried and re-suspended in 1 mL of dimethylsulfoxide (approximately 100 mg/mL). They were then diluted 1 to 40 into DMSO. This 1 to 40 solution was diluted 1 to 25 into 50:50 (MEOH:H_2O). This 1 to 1000 solution was diluted 1 to 20 into 50:50 (MEOH:H_2O) for a final dilution factor of 20,000 (approximately 5 µg/mL).

LC Conditions

LC system: ACQUITY UPLC I-Class with FTN Sample Manager
Column: ACQUITY UPLC BEH 2.1 x 50 mm, 1.8 µm, 50 °C
Sample temp.: 10 °C
Mobile Phase: A: water (0.1% FA); B: acetonitrile (0.1% FA)
Flow Rate: 0.8 mL/min
Gradient:

Time (min)	% A	%B	Curve
Initial	95	5	Initial
0.2	95	5	6
3.2	5	95	6
3.8	0	100	1
4.5	95	5	1

MS Conditions

MS system: Xevo G2-XS QTof MS
Acquisition range: 150-1800 Da (0.1 s scan rate)
Acquisition mode: MS^E, ESI^- and ESI^+ in resolution mode
Capillary voltage: 3 kV (ESI^+)/2.5 kV (ESI^-)
Cone voltage: 30 V
Collision energy (eV): Low CE: 6; High CE: 25-45
Source temp.: 120 °C
Desolvation temp.: 500 °C

Data Processing Platform

The data processing platform integrates image-based phenotypic profiling data from our recently reported cytological profiling platform with untargeted metabolomics data from UPLC/QTof MS platform. MS data was collected using data independent acquisition (MS^E) which simultaneously provides exact mass precursor and corresponding fragment ions for identification and structural elucidation. Using a custom informatics platform, image-based phenotypic and UPLC/QTof MS datasets are integrated to identify candidate molecules that are consistently positively correlated with specific phenotypes. Using network display, the bioactive metabolome from the natural product library is then displayed as an annotated network diagram that identifies all sets of bioactive molecules from within this set, allowing the selection and development of high priority lead compounds.

RESULTS AND DISCUSSIONS

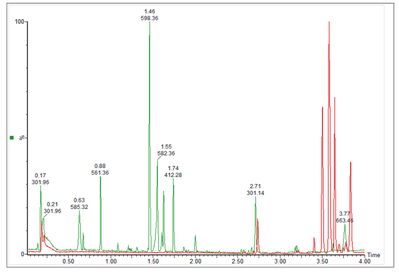

Figure 2. Representative overplayed chromatograms of the early and late eluting different fractions in positive ion mode.

One of the bottlenecks in marine natural product research is that, the absence of a custom marine microbial compounds library with precursor exact mass, fragment ion information, retention time, collision cross section area (CCS) and theoretical isotopic distribution that allows for confident identification of compounds from a complex microbial sample extracts. We are in the process of building a custom marine microbial compounds library which contains compound name, chemical structure, molecular formula, average molecular mass and accurate mono-isotopic molecular mass of each compound. Figure 3 shows the basic infra structure of the library and the information it contains. More over as indicated in the red insert rectangle, the library can be populated with information such as compound properties (physical properties, specific identifiers and synonyms), literature references, documents, experimental spectra etc.

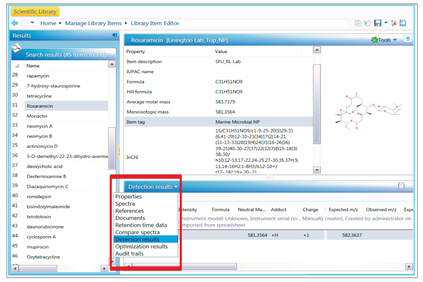

Figure 3. The basic infra structure of the marine natural products library in UNIFI.

RESULTS AND DISCUSSIONS

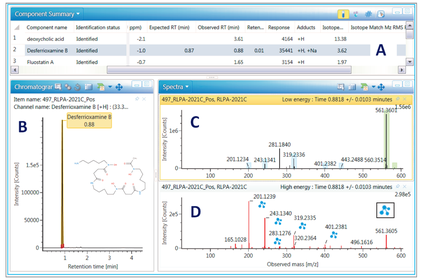

Figure 4. Identification result from a custom marine microbial library. (A) The component summary interface; (B) Selected ion chromatogram of single component corresponding to panel A (C) The respective low energy precursor exact mass spectrum and (D) The corresponding high energy fragment ion spectrum. In the high energy MS^E spectrum, the blue " " mark indicates the experimental fragment ion that matches to the expected in silico fragment ions generated from the mol structure using MassFragment.

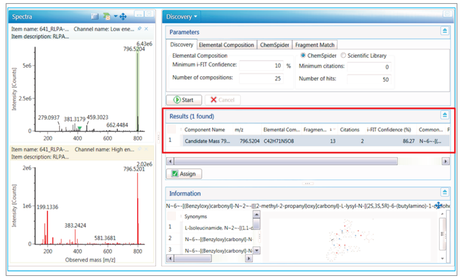

Figure 5. Structural elucidation for the identification and confirmation of the unknown components.

Structural elucidation tools in UNIFI was used for the identification of unknown high intensity peaks. Batch search can be performed using online databases such as Chemspider. The key steps for the unknown structural elucidation are (Figure 5):

1. Set the basic search parameters such as possible elemental composition, fragment match and isotopic distribution in the discover.
2. Search against Chemspider database (on-line) containing about 600 libraries.
3. The initial match was validated and confirmed by analyzing fragment ions obtained in MS^E scan via Embedded *in-silico* tool MassFragment™.
4. After final identification results are reviewed and confirmed, the identified compounds can be sent to the scientific library with all information including exact mass, retention time and fragment ion.

RESULTS AND DISCUSSIONS

Finally, in-house informatics platform developed at SFU was used to predict the biological activities of constituents of bioactive mixtures, and use these data to refine predictions about which constituents are responsible for the observed phenotypes in image-based whole cell screening data. This combination of multimode high-resolution untargeted metabolomic profiling and multiparametric biological annotation provides an opportunity to highlight even very minor bioactive constituents from natural product extracts that cause strong and reproducible phenotypes (Figure 6).

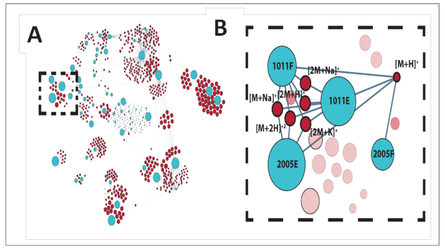

*Figure 6. Illustration of how Compound Activity Mapping enables discovery. **(A)** Network displaying m/z features associated with consistent bioactivity. **(B)** Zoom in of the staurosporine cluster with extracts and relevant adducts labeled.*

CONCLUSION

- **The Natural Products Application Solution with UNIFI provides a single workflow for data acquisition, processing and confident compound identification based on low energy precursor exact mass, theoretical isotopic distribution and corresponding high energy fragment ion information from custom marine microbial scientific library or Chemspider.**
- **Integrating image-based screening and high-resolution UPLC/QTof MS provides a comprehensive annotation of the identities and biological attributes of all bioactive constituents.**
- **This technology provides natural products chemists with a new mechanism to convert complex metabolomics profiling of extract libraries into a Compound Activity Map, which clusters extracts and metabolites based on common chemical and biological properties and highlights those compounds predicted to be responsible for the observed phenotype of a particular extract.**

References

1. Kurita L. K. Linington R. G.; "Connecting Phenotype and Chemotype: High-Content Disovery Strategies for Natural Product Research"; J. Nat. Prod. 2015, 78, 587-596.
2. Using Natural Products Application Solution with UNIFI for the Identification of Chemical Ingredients of Green Tea Extract. Waters. 2013; 720004837en.

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS